Risk Premia and Option Intermediation[*]

Thomas Gruenthaler[‡]

Latest version here

Abstract

Equity index risk premia vary more than can be explained by market risks and pricing models. I show that index option intermediaries cause variation in risk premia to manage their option positions. When expected volatility is low, intermediaries hold risky short positions. Increasing risk and liquidity premia compensate intermediaries for the risk exposure and lower subsequent demand. When expected volatility is high, intermediaries transfer risky short positions to investors. Intermediaries induce sell orders by quoting relatively higher sell prices and charging wider effective spreads for buyer-initiated trades. As a result, intermediaries' positions are often riskless when high volatility realizes. Tighter constraints in the financial sector transmit to option intermediaries and create a reluctance to hold positions. The reluctance drives level and variation in crash risk premia independent of market risks. My results suggest that financial institutions' risk tolerance integrates option markets with other markets.

Keywords: Risk Premia, Intermediary Asset Pricing, Option Returns, Option Pricing, Crash Risk

JEL: G12, G13

[*] I am deeply indebted to Torben Andersen, Nicole Branger, Christoph Schneider, and Viktor Todorov for their advice and helpful comments. I am also grateful for very helpful comments from Heiner Beckmeyer, Jeroen Dalderop, Leander Gayda, Jan Harren, Ariel Lanza, Friedrich Lorenz, Paul Meyerhof, Laura Rettig, Timo Wiedemann and particpants at the Quantitative Finance Seminar at Kellog School of Management, Seminar at University of Münster, Tilburg University.

[‡] Finance Department, Tilburg University, The Netherlands. E-mail: t.grunthaler@tilburguniversity.edu

I. INTRODUCTION

A key feature of many markets is that financial institutions serve as intermediaries and provide liquidity by taking the opposite side of a trade. Shocks to the capital of financial institutions decrease the ability to provide liquidity and increase risk premia (Kondor and Vayanos, 2019). Consequently, variations in risk premia of various asset classes are related to the 'health' of the financial sector (Adrian, Etula, and Muir, 2014; He, Kelly, and Manela, 2017). Meanwhile, premia for higher-moment risks such as volatility or crashes are hard to reconcile with asset pricing models.[1] Why are higher-moment risk premia so large? Why do they sometimes vary, although risks do not? Why does the crash risk premium stay elevated after turbulent market episodes, although risks revert to normal?

This paper shows that index option intermediaries (OptInt) influence risk and liquidity premia to manage the risks associated with their option positions. First, I show that the level of risk and liquidity premia increases when OptInts' positions become riskier. Second, OptInts manage to transfer risky short positions to non-intermediaries when expected volatility is high, although OptInts cannot directly control their positions. OptInts increase sell prices more than buy prices and charge higher spreads for buyer-initiated trades. The asymmetries induce sell orders, and, as a result, the risks of extreme losses for OptInts are shallow around market crashes. Third, tighter constraints in the financial sector affect risk-bearing capacities, and OptInts become reluctant to hold positions. The reluctance is reflected in the level of the crash risk premium and induces variation independent of volatility. The results suggest that financial institutions integrate option markets with other markets.[2] Decreasing risk tolerance transmits directly to option prices and ultimately affects other asset classes through contractions of the balance sheets of financial institutions (Adrian and Shin, 2010, 2014).

Option markets are well suited to analyze the interplay between financial intermediation and risk premia. Options come with different strikes and maturities, loading differently on various sources of risk. Hence, options are uniquely well suited for extracting higher-moment risk expectations and premia. Options are zero-net-supply assets with two counterparties holding offsetting positions. The market is intermediated by dealers such as Goldman Sachs that provide depth and liquidity by absorbing order imbalances. When mutual funds buy options but there is no seller, OptInts execute the trade by writing a new option contract. Therefore, higher demand results in rising prices (Gârleanu, Pedersen, and Poteshman, 2009). OptInts hold 50% of the total outstanding value of index options. The positions bear exposure to extreme losses because options cannot be hedged entirely (Muravyev, 2016). The requirement of OptInts to provide liquidity has two consequences: First, OptInts accumulate positions and risks over time. Second,

[1] See Andersen, Fusari, and Todorov (2015b) or Beason and Schreindorfer (2022).

[2] Zhou (2018) documents that variance risk premium predicts returns of equities, bonds, currencies, and CDS.

OptInts cannot directly manage the risk of their positions. They absorb order imbalances regardless of their desire to hold specific options. The tool to control the risk of OptInts' positions is adjusting option quotes and spreads, encouraging investors to trade in a particular direction.

To measure the risk of OptInts' positions, I propose an ex-ante option intermediary risk (*OIR*) factor that is motivated by the risk-management systems of OptInts (Bergomi, 2016). I quantify the worst-case profit and loss (P&L) of OptInts' positions by stress test scenarios. *OIR* quantifies the P&L of option positions conditional on a hypothetical market crash defined as an extreme negative stock market drop accompanied by rising volatility. The measure captures a crash event's impact on OptInts' P&L, not the probability of a crash. Negative levels of *OIR* indicate that OptInts are net short and bear the risk of capital losses in case of a crash. Positive levels indicate that OptInts have hedged net long positions that profit from stock market crashes. The factor captures the risk of the entire option intermediary sector, incorporates the total exposure to market risks, is independent of the motive behind investors' option trades (economic fundamentals, asymmetric information), and assesses the effect of every transaction relative to OptInts' current risk exposure.

For my empirical analysis of the impact of OIR on premia, I use high-frequency option quotes and calculate hedged option returns. Hedged option returns proxy the risk premia inherent in options (Bakshi and Kapadia, 2003). Through panel regressions and lag identification, I establish that a one standard deviation decrease in *OIR* generates an expected return of 0.5% overnight, 1.34% the next trading day, and 1.74% day-to-day. The effect goes beyond volatility risk, index returns, and order imbalances. The coefficient of lagged *OIR* is significantly larger than that of unlagged order imbalances, suggesting that OptInts assess the impact of option trades not in isolation but with regard to aggregate positions. I document a similar relationship for liquidity premia in option markets. I use realized S&P 500 option transactions and show that OptInts charge higher effective spreads for providing liquidity when *OIR* is more negative (OptInts have more short positions). Rising risk premia and widening spreads control aggregate positions and OptInts' income.[3] If investors start selling options to earn the increased premia, the risk of extreme losses will subsequently reduce. However, OptInts have to pay a higher premium and, thus, widen spreads. If investors continue buying, the risk of extreme losses will further increase. OptInts receive higher premia and spreads as compensation for bearing the risks.

Reduced form correlations do not conclusively establish causality between *OIR* and premia. *OIR* responds to changes in quantities and prices, and so might tomorrow's option premia. I use three sets of instrumental variables to overcome the potential endogeneity problem. The first is past order imbalances. Chordia and Subrahmanyam (2004) show that past order imbalances predict future variation in order imbalances and, hence, in OptInts' aggregate positions. The

[3] See Amihud and Mendelson (1980) and Ho and Stoll (1981).

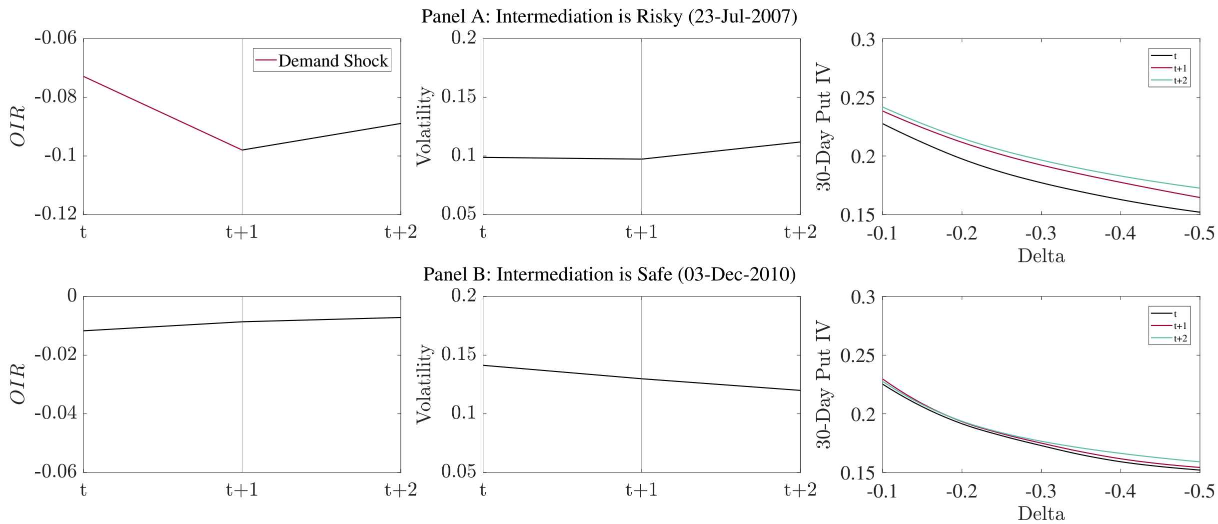
second instrument is a dummy for option expiry days (Muravyev, 2016). The third instrument is a dummy that indicates tighter constraints in the financial sector. I use shocks to the leverage ratio of primary dealers, the largest financial institutions, and the most relevant option intermediaries. Intuitively, a decrease in equity capital impairs the willingness to supply options (Chen, Joslin, and Ni, 2019). The instrumental variable estimates reinforce a negative causal effect of OIR on premia. I find that a one standard deviation decrease in instrumented OIR is associated with a daily hedged option return of 0.51%, with half of the return earned overnight and half intraday. Strikingly, the coefficient of instrumented OIR is almost unchanged throughout a series of robustness checks. I test different constructions of the factor, exclude high volatility episodes, and control for the realized P&L of OptInts. The associated return ranges from 0.48% to 0.56%.

To illustrate the effect of OIR, I show two days with different OIR levels but identical implied volatility curves in Figure 1. I express option prices in implied volatilities (IV) to ensure comparability across time and option characteristics. Panel A highlights that option prices boost after a shock to OIR on July 20$^{\text{th}}$, 2007. The upper left plot shows that OIR decreased by almost 35% from t to $t+1$, suggesting that OptInts would have more extreme capital losses if markets crashed. IVs increased by almost 10% from t to $t+1$, although volatility stayed flat. Options became substantially more expensive, and investors started selling to receive the premia. This boosted OIR at $t+2$. Once volatility increased ($t+2$), IVs did not change much because OIR was already recovering. Panel B shows a day in Dec 2010 with an identical volatility curve. OIR is near zero, i.e., the risk of extreme losses is low, and volatility decreases. Therefore, OptInts did not need to incentivize specific trading; hence, option prices remained almost unchanged.

The price increase in Panel A is likely a product of a more severe impact of a hypothetical crash event, captured by OIR, and distress in the financial sector. Two funds of Bear Stearns filed for bankruptcy protection on July 25$^{\text{th}}$, 2007. My interpretation of these findings is that distress in the financial sector impairs risk tolerance and transmits to OptInts. Brunnermeier and Pedersen (2009) show that financial institutions are reluctant to hold positions during episodes of tight funding constraints. Adrian and Shin (2010) argue that the risk tolerance of the financial system is reflected in asset prices because financial institutions expand or shrink balance sheets.

The previous findings motivate my subsequent analysis of the interplay between the risk tolerance of the financial sector and OIR. I document that high leverage, a bad economic outlook, tight funding liquidity, and increased market volatility correlate strongly with lower risk exposure of OptInts. In fact, I find that such episodes are characterized by OIR being close to zero or positive, indicating that OptInts would only make small losses or profits when markets crashed. OptInts' risk-bearing capacities reduce in anticipation of higher market volatility or when constraints in the financial sector tighten, resulting in a reluctance to hold risky positions. The results imply that OptInts' positions turn from net short to net long. The evidence

Figure 1. Option Intermediary Risk and Volatility Surface



The figure shows the dynamics of absolute OIR, realized volatility, and option prices (in implied volatilities) for two days with identical volatility surface. Panel A shows the dynamics when OIR is low (t = Jul 20, 2007) and Panel B when OIR is high (t = Dec 03, 2010). OIR denotes intermediaries' additional capital losses per option contract in case of a market crash, defined in Section III. Volatility is the one-week average of daily 5-min realized volatilities from the Oxford-Man Realized Library. Put IVs are taken from OptionMetrics Volatility Surface.

complements Chen, Joslin, and Ni (2019), who show that OptInts buy out-of-the-money puts when constraints are tight. However, OptInts cannot simply buy options because they absorb the imbalances of other investors' trading.[4]

How can OptInts influence investors to sell options when market volatility is high or the risk tolerance in the financial sector is low? I show that OptInts actively incentivize investors to take over undesired short positions. The unconditional coefficients of regressing bid and ask option returns on OIR are identical but diverge when market volatility is high. The difference equals 60% in the instrumental variable approach. I find a similar asymmetry for various option return predictors, such as the VIX and index returns. I also document that effective spreads for buyer-initiated transactions are higher than for seller-initiated. Sell (bid) prices rise more than buy (ask) prices, and transaction costs for buying are higher than for selling. The asymmetries induce selling activity and explain how OptInts reduce short positions.

In the last part of the paper, I estimate an option pricing model to study the relationship between crash risk premia and OptInts' positions. I parametrize the model such that high premia coincide with low OIR because these are episodes in which OptInts anticipate high crash risk or the risk tolerance of the financial sector is low. The model estimation provides important

[4]An individual OptInt might trade options with another OptInt, but the trade is just a transfer within the intermediary sector. OptInts require other investors, such as hedge funds or mutual funds, as counterparts.

implications for option pricing and the puzzling behavior of risk premia. On average, two-thirds of the crash probability is driven by OIR and only one-third by market volatility. OIR's contribution is low when financial institutions have a higher risk tolerance, e.g., before the financial crisis, but sharply increases before markets become volatile. This indicates that OptInts anticipate turbulent episodes and increase crash risk premia to incentivize option selling. Hence, *OIR* induces variation in crash risk premia independent of volatility (Bollerslev and Todorov, 2011).

I also document that the contribution of *OIR* to the crash probability remains persistently elevated after market crashes. OptInts are reluctant to hold option positions because the financial sector's risk tolerance is low. The tolerance decreases because of incurred losses in other assets, higher regulatory constraints, less capital, and fewer intermediation capacities (bankruptcies). This affects financial institutions and ultimately transmits to OptInts through tighter risk limits. Option prices reflect the reluctance, which is why crash risk premia stay elevated after crashes although volatility is low (Bates, 2000; Jackwerth, 2000; Andersen, Fusari, and Todorov, 2020).

The results suggest that financial institutions integrate markets because prices reflect their risk tolerance. Option-implied risk premia predict returns of various asset classes because all prices respond to shrinking or expanding balance sheets of financial institutions (Adrian and Shin, 2010). Hence, crash risk premia are partly self-generated by the financial system and do not only reflect fundamentals (Brunnermeier and Sannikov, 2014). Stronger regulatory policies are designed to make the financial sector more resilient. However, they could also tighten constraints, thereby increasing risk premia and liquidity costs in the option market.

Literature Review: My paper relates to a growing literature studying how intermediaries impact risk premia (Brunnermeier and Pedersen, 2009; He and Krishnamurthy, 2013; Brunnermeier and Sannikov, 2014). Adrian, Etula, and Muir (2014), He, Kelly, and Manela (2017), Muir (2017), Haddad and Muir (2021), and Du, Hébert, and Huber (2022) provide empirical evidence that relates intermediaries to returns of several asset classes. Chen, Joslin, and Ni (2019) infer intermediary constraints from traded out-of-the-money put volumes and show that months in which other investors net sell puts to intermediaries predict risk premia for several asset classes. Cheng (2019) shows that the volatility risk premium falls or stays flat when ex-ante measures of volatility rise and relates the findings to decreasing demand for volatility hedges. Both results seem unintuitive because they imply that investors buy options when volatility is low but sell options when volatility is high. My contribution is to show how OptInts incentivize this pattern. They induce public selling by quoting higher bid prices and charging higher effective spreads for buyer-initiated trades.

Another strand of literature relates option prices to demand effects. Bollen and Whaley (2004) show that demand impacts option prices. Gârleanu, Pedersen, and Poteshman (2009) derive an equilibrium model in which demand pressure enters the pricing kernel. Muravyev (2016) establishes that order imbalances have a more substantial effect on equity option prices than asymmetric information and highlights the predictive power of past order imbalances. In contrast to these papers, I focus on the dynamics of aggregate index option positions and show that option intermediary risks impact prices beyond demand pressure. Barras and Aytek (2016) show that the difference between the variance risk premium inferred from equity and option markets is related to the financial standing of intermediaries. Almeida and Freire (2022) find that demand for options helps to explain the puzzling shape of the pricing kernel. Fournier and Jacobs (2020) derive a structural model in which the variance risk premium is a function of intermediaries' wealth and volatility risk. In contrast, I find that episodes with low option intermediary risk are more informative because they reflect intermediaries' crash risk expectations and risk tolerance.

II. DATA AND INTERMEDIARY POSITIONS

This section describes the original datasets, the processing of the data, and the aggregate intermediary positions. Appendix A provides additional details on the data.

A. Data Description

Volume: S&P 500 index options exclusively trade on the CBOE. The Open-Close Volume Summary database offers detailed information on trading activity. I work with the end-of-day volume summary of all options traded on the primary exchange C1. The data is available from January 1, 1990, until November 30, 2020, and aggregates transactions for every traded option by trade origin (customer/firm), type of transaction (buy/sell), and type of position (opening/closing). Investor trading via a broker exemplifies customer-originated trades. In contrast, an option trader who trades for the broker's account exemplifies a firm-originated trade.[5] An open transaction opens a new option position, while a close transaction closes an existing position. It is essential to distinguish between the two types. The former increases the number of outstanding contracts while the latter reduces it. The data does not provide information on the specific option intermediary. Therefore, OptInts are the aggregate, representative option intermediary.

End-of-Day Option Quotes: The volume data comes without information on prices. I use end-of-day prices from OptionMetrics for the period from January 1996 until the end of the sample. For the period before 1996, I obtain quotes from LiveVol. The data reports open interest, volume, and bid/ask prices. I use the midpoint between bid and ask prices to compute Black-Scholes greeks. The risk-free rates are maturity adjusted by linearly interpolating the Treasury curve.

[5] As of 2011, the data is more granular and reports more subgroups for the trade origin. Jacobs, Mai, and Pederzoli (2021) show that this has little impact on the classification of trades.

Intraday Option Quotes: I use intraday option quotes from LiveVol to determine delta-hedged option returns. The data comes at a 1-minute frequency and is available from January 2004 until the end of 2020. I use filters along the lines of Andersen, Bondarenko, and Gonzalez-Perez (2015) and Muravyev and Ni (2020).

Option Trades: LiveVol provides intraday option transactions. The data is available from January 2004 until the end of 2020, and the timestamp is in milliseconds. The data provides information about the traded volume, the trade price, implied volatility, and bid/ask prices. I apply the filters from Andersen et al. (2021).

B. Data Processing

Previous studies that have worked with the Open-Close Volume Summary apply specific filters to the data. I think it is essential to consider all transactions to provide a complete picture. OptInts have to hold and hedge all types of options, regardless of maturity, moneyness, or implied volatility. Therefore, I generally do not filter the volume data and keep all maturity and moneyness categories available. The only exception is the exclusion of so-called FLEX Options with customized contract terms, such as strike prices in penny increments, for which no price information is available. The filter kicks out approximately 0.2% of all option transactions.

I experienced two significant sources of error working with the Open-Close Volume Summary database. First, for the years in which the S&P 500 was around 1000, the data cuts 1000 from four-digit strikes. For instance, the data reports a strike price of 1025 as 25. Second, some option chains come with a negative time-to-maturity. Appendix A describes a simple method for correcting the data and shows how essential the correction is. Up to 90% of the data in some months would have been dismissed otherwise. For every day in the sample, I match the volume data with option price information using an identifier that consists of expiry, strike, and put-call flags. After initially matching each option trade with the quote on the trading day, the next challenge lies in assigning updated prices (or implied volatilities) over the period in which OptInts hold the option. All options with outstanding open interest should be updated daily, but this is not the case. For any day a quote is unavailable, I fill the missing implied volatility by using interpolated implied volatilities (Jiang and Tian, 2005).

C. Aggregate Intermediary Positions

OptInts try to match buys with sales so that the spread income is largest and transactions have no impact on aggregate positions. However, the immediacy service also consists of taking the opposite side of a trade in case buys and sells do not arrive synchronously. OptInts' new positions are the residual of the sum of customer- and firm-originated trades. The residual is often referred to as order imbalance. I follow Ni, Pearson, Poteshman, and White (2021) in calculating the aggregate positions. For each day t and option series j, I determine the buy and sell open interest OI as

$$OI^{\text{buy}}_{j,t} = \underbrace{OI^{\text{buy}}_{j,t-1}}_{\text{Existing}} + \underbrace{Volume^{\text{OB}}_{j,t} - Volume^{\text{CS}}_{j,t}}_{\text{Order Imbalance}}, \tag{1}$$

$$OI^{\text{sell}}_{j,t} = OI^{\text{sell}}_{j,t-1} + Volume^{\text{OS}}_{j,t} - Volume^{\text{CB}}_{j,t}, \tag{2}$$

where $Volume^{\text{OB},y}_{j,t}$ are opening buy transactions and $Volume^{\text{OS},y}_{j,t}$ are opening short transactions. $Volume^{\text{CS},y}_{j,t}$ are transactions that close existing long positions and $Volume^{\text{CB},y}_{j,t}$ are transactions that close existing short positions. Both decrease the number of outstanding contracts because the buyer/seller now owns an offsetting position. The positions are determined from all non-market maker trades. OptInts have to take the opposite position so that the aggregate position in option j is given by

$$Pos_{j,t} = -1 \cdot \left(OI^{buy}_{j,t} - OI^{sell}_{j,t} \right). \tag{3}$$

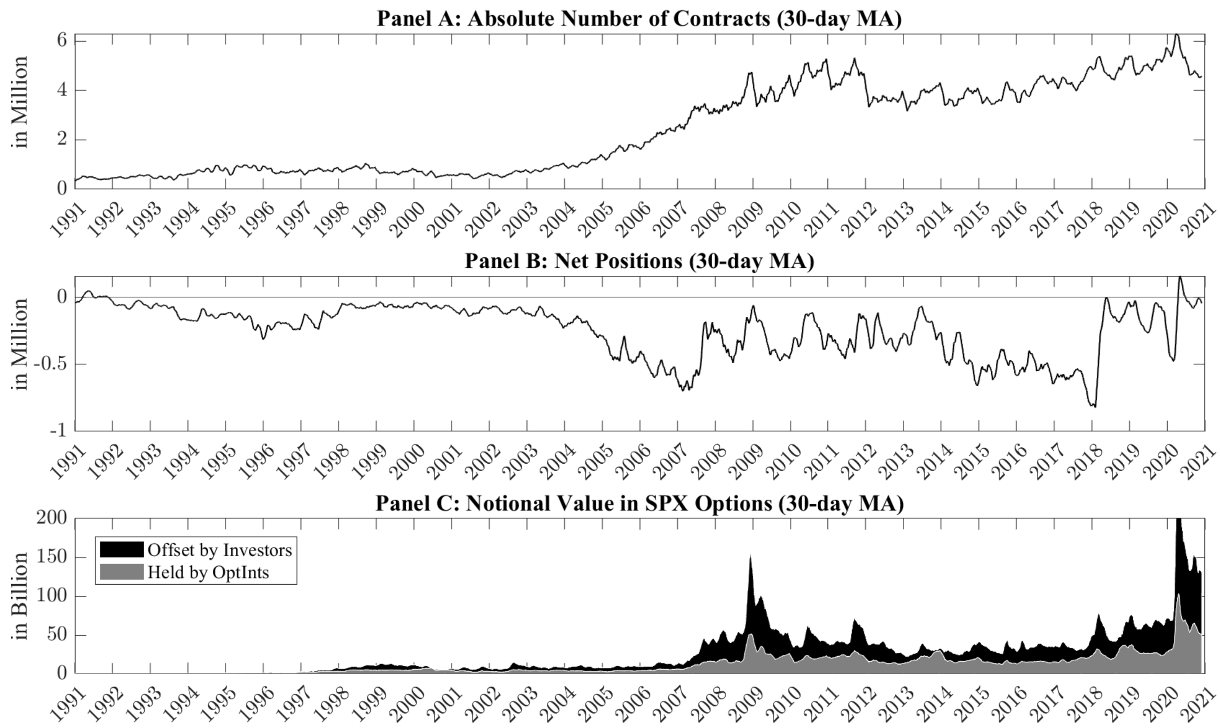
OptInts have a long position in option j when accumulated sell order volumes exceed buy order volumes and vice versa. Figure 2 provides some intuition about OptInts' positions.[6] Panel A shows the absolute number of contracts intermediaries hold. OptInts have managed up to one million contracts during the 90s. The market has substantially grown since 2003, and OptInts manage up to six million option contracts as of 2020. Panel B shows that OptInts are almost always aggregate net sellers. Put options drive most of the net position (see Figure G.2 in the Appendix). Panel C plots the total market notional and the share held by OptInts. The notional has also grown significantly, with more than $100 billion outstanding after the recent retail boom. Most importantly, the plot shows that OptInts have a large stake in the whole market and are, on average, invested in more than 50% of the market's value. That suggests that OptInts have to accommodate a huge fraction of order imbalances.

III. OPTION INTERMEDIARY RISK

This section introduces option P&Ls, discusses the regulatory framework for option risk management, and describes the final option intermediation risk measure.

[6] I use 1990 as the burn-in period.

Figure 2. Intermediary Positions in Index Options



The figure shows the option intermediaries' position from 1991 until 2021. Panel A plots the absolute number of outstanding contracts defined as $|OI_{j,t}^{buy} + OI_{j,t}^{sell}|$, Panel B the net position defined as $-1 \cdot (OI_{j,t}^{buy} + OI_{j,t}^{sell})$, and the notional value defined as the absolute number of outstanding contracts scaled by the option value and the contract multiplier of 100.

A. P&L Attribution

Even after ignoring issues such as model misspecification, parameter uncertainty, discrete frequencies, and transaction costs, quantifying the risk of OptInts' positions is a complex endeavor. Most factors center around raw or risk-weighted order imbalances but ignore the effect of aggregate positions. I propose a factor based on the P&L of all intermediary positions. An option's price usually depends on inputs such as the underlying S, the implied volatility IV_t, and time t. Because derivative instruments are ultimately just functions of their inputs, Taylor series expansion can be used to attribute changes in option prices to different risk contributions. A pricing equation (PDE) must be formulated to disentangle the P&L into different sources of risk. Note that the PDE merely reflects a tractable accounting tool. The literature and derivative desks almost always use the BSM pricing equation for the expansion (Sepp, 2012; Bergomi, 2016; Carr and Wu, 2020). The choice of the pricing equation does not imply that the same model has been used to price the option. Implied volatilities are an intuitive illustration of this. No derivative desk uses the Black and Scholes (1973) model (BSM) for pricing, but every desk uses the equation to translate the prices into implied volatilities. Using the BSM equation implies that the implied volatility captures all shocks other than in the underlying.

I follow this practice and use the BSM equation and further assume that dividends are paid continuously so that the forward value of the underlying is $F_t^T = Se^{(r-q)(T-t)}$. The instantaneous change in a long call $C_t(F_t, K, IV_t, \tau)$ is given by

$$\begin{aligned} dC_t = &\frac{\partial C_t}{\partial F_t} dF_t + \frac{\partial C_t}{\partial t} dt + \frac{\partial C_t}{\partial IV_t} dIV_t + \frac{1}{2}\frac{\partial^2 C_t}{\partial F_t^2}(dF_t)^2 \\ &+ \frac{1}{2}\frac{\partial^2 C_t}{\partial IV_t^2}(dIV_t)^2 + \frac{\partial^2 C_t}{\partial S_t \partial IV_t}(dF_t\, dIV_t). \end{aligned} \tag{4}$$

The option price is expanded to the first order in price (*Delta*, Δ), time (*Theta*, Θ), implied volatility (*Vega*, v), and second order in price (*Gamma*, Γ), implied volatility ($Volga$, ϑ), and their product ($Vanna$, ε). The expansion up to the second derivative suffices to bring the approximation error to an order lower than the investment horizon (Bergomi, 2016; Carr and Wu, 2020). The partial derivatives are well-known option greeks that measure the sensitivity of the option price with respect to its inputs. Δ (v) is the directional exposure to the underlying (implied volatility). Γ (ϑ) measures how Δ (v) changes when the underlying (implied volatility) moves. ε is the change in Δ when implied volatility moves. Θ quantifies the time decay of an option. Θ is negative because, as time passes, the option is less likely to be in the money. Carr and Wu (2020) use this result and derive that the fair value of an option must be set so that the other sensitivities balance out the loss in time value. This must hold under no-arbitrage and risk-neutral expectations.

I use this result but consider realizations. Intermediaries hedge against the directional exposure implies that the dF_t term in Equation (4) drops. Expressing all partial derivatives in percentage units (cash sensitivities) instead of units of the underlying and dividing by the investment horizon yields

$$\frac{dC_t}{dt} = \Theta + \frac{1}{2}\Gamma F_t^2 \sigma_t^2 + v IV_t \mu_t + \frac{1}{2}\vartheta IV_t^2 \omega_t^2 + \varepsilon\, F_t\, IV_t \gamma_t. \tag{5}$$

where

$$\sigma_t^2 = \left(\frac{dF_t}{F_t}\right)^2 /dt, \quad \mu_t = \left(\frac{dIV_t}{IV_t}\right)/dt, \quad \omega_t^2 = \left(\frac{dIV_t}{IV_t}\right)^2 /dt, \quad \gamma_t = \left(\frac{dF_t\, dIV_t}{F_t\, IV_t}\right)/dt, \tag{6}$$

are the realized variance of the underlying, the change in IV, the realized variance of implied volatility, and the covariance rate of underlying and implied volatility. Conveniently, all greeks can be expressed in terms of ΓS^2. That is,

$$\Theta = -\frac{1}{2}\Gamma F_t^2 IV_t^2, \quad v\, IV_t = \tau IV_t^2 \Gamma F_t^2, \quad \vartheta\, IV_t^2 = \Gamma F_t^2 z^- z^+, \quad IV_t F_t \varepsilon = \Gamma F_t^2 z^+, \tag{7}$$

where τ is the time-to-maturity and $z^\pm = \ln\left(\frac{K}{S}\right) \pm \frac{1}{2} IV^2 \tau$ is a convexity-adjusted measure of moneyness.[7] Plugging in and rearranging gives the final P&L normalized by the investment horizon:

$$\frac{dC_t}{dt} = \frac{1}{2}\Gamma F_t^2 \left(\sigma_t^2 - IV_t^2 + 2\tau IV_t^2 \mu_t + z^- z^+ \omega_t^2 + 2z^+ \gamma_t\right). \tag{8}$$

The decomposition shows that OptInts are prone to several sources of risk that cannot be delta-hedged. The term is the contribution that comes from the difference between realized and implied variance. The second is the change in the implied variance scaled by the current level and maturity of the option. The third term is the variance of implied volatility scaled by the option's moneyness. The last term captures the correlation between changes in the underlying and changes in implied volatility. The respective contributions to the P&L depend on many factors, such as maturity and moneyness. Short-term ATM calls load primarily on the difference between implied and realized variance, while OTM/ITM calls load on the variance difference and the covariance. Long-term calls load strongly on the variance difference and the change in implied volatility. To gain more intuition, I describe the P&L for two specific cases.

Constant Implied Volatility: To highlight that providing immediacy for the option market is risky even under the most simplifying circumstances, I assume that implied volatilities are constant. All option sensitivities concerning implied volatility disappear, and the P&L of a long option reads

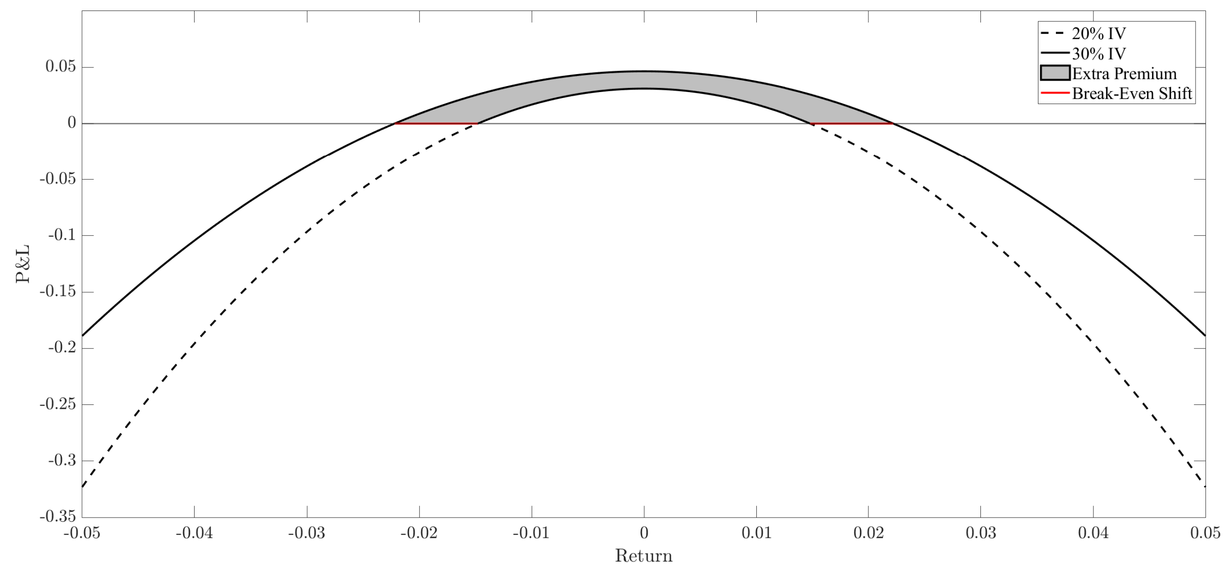
$$\frac{dC_t}{dt} = \frac{1}{2}\Gamma F_t^2 \left(\sigma_t^2 - IV^2\right). \tag{9}$$

The decomposition is visualized in Figure 3 and shows that a delta-hedged short option is a non-linear bet against high variance. Selling an option pays off when realized variance is lower than implied, which is attributable to the option payoff's non-linearity and the delta hedge's linearity. The premium for the non-linear feature is the option's loss in value over time (theta premium). Higher implied volatilities increase the theta premium and shift the break-even point of the P&L. Higher implied volatilities are favorable for option sellers. Hence, from the perspective of an intermediary predominantly selling options, higher implied volatilities increase the probability of making profits and decrease the probability of making losses.

Stochastic Volatility and Jumps: Models that feature stochastic volatility and jumps are

[7] See Hull (2009) and Carr and Wu (2020).

Figure 3. Daily P&L of a Short Delta-Hedged ATM Call



The figure shows the daily P&L at $t + 1$ of a short delta-hedged call option as a function of the underlying's return. The solid black line shows an option with 20% and the dashed black line with 30% implied volatility. The underlying F at t is 100, strike K is 100, and time-to-maturity τ is 0.5 years. The risk-free rate and dividend yield are set to zero.

usually specified as

$$\begin{aligned} dF_t &= \alpha F_t dt + \sqrt{V_t} F_t dW_t + (e^x - 1)\tilde{\mu}(dt, dx), \\ dV_t &= \kappa(\overline{V} - V_t)dt + \sigma_v \sqrt{V_t} dB_t, \end{aligned} \tag{10}$$

where α is the drift of the forward underlying, κ is the mean-reversion speed of variance, $\overline{V}$ is the mean-reversion level, σ_v the volatility of volatility, and dB are Brownian motion that has a correlation of ρ to the Brownian motion of the asset price dW. The measure $\tilde{\mu}$ counts the number of jumps with intensity λ_t and magnitude x. The model complicates the P&L because of the autocorrelation of variance, and not all partial derivatives are available in closed form.[8] However, both effects do not change the overall interpretation. I assume that OptInts use BSM sensitivities as ad-hoc choices. The instantaneous P&L can be approximated by

$$\frac{dC_t}{dt} \approx \frac{1}{2}\Gamma F_t^2 \left(\sigma_t^2 - \Sigma_t + 2\tau\Sigma\mu_t + z^- z^+ \omega_t^2 + 2z^+\gamma_t\right). \tag{11}$$

[8]Sepp (2013) shows the P&L with the autocorrelation correction. The derivation is slightly less complex because it underlies the PDE of the BSM that does not account for volatility sensitivities.

where the model's implied variance is

$$\Sigma_t = \overline{V} + \frac{1}{\kappa\tau}\left(1 - e^{-\kappa\tau}\right)(V_t - \overline{V}) + \lambda_t \mu_j, \tag{12}$$

and

$$\mu_t = \frac{\kappa(\overline{V} - V_t)}{\Sigma_t}, \quad \omega_t = \frac{\sigma_v^2 V_t}{\Sigma_t}, \quad \gamma_t = \frac{\rho\sigma_v V_t F_t}{\Sigma_t F_t} = \frac{\rho\sigma_v V_t}{\Sigma_t}. \tag{13}$$

The approximate P&L gives a clearer picture of each component's contribution. Consider the results from the perspective of an intermediary who sells a call option. There is a constant risk that volatility realizations are higher than expected, e.g., because a jump realizes ($\int_0^T \int_{\mathbb{R}} x^2 \mu(ds, dx) > 0$) or volatility suddenly shifts. Of course, the implied volatility of options is greater to make the break-even point of the P&L wider. Intermediaries also make losses when the current variance level is below its long-term mean $\overline{V} > V_t$, which is scaled by κ. Lower values for the mean-reversion speed mitigate the impact of this risk. This is exactly the effect of the volatility risk premium in standard option pricing models, which lowers the mean-reversion speed under the risk-neutral probability measure (Bates, 2003). The contribution of the variance of implied variances is ambiguous and depends on the option's moneyness. While ω_t is strictly positive, $z^{\pm}$ determine if the intermediary profits from fluctuations. Generally speaking, $z^{\pm}$ are positive for OTM calls, negative for ITM calls, and have opposite signs for ATM calls. The last contribution comes from the covariance rate of the underlying and implied volatilities. γ_t is negative, and z^+ is positive for OTM and ITM calls and negative for ITM calls.

B. Regulatory Requirements

The risks associated with option trading are hard to manage and often result in extreme losses. In 2017, Catalyst Capital Advisors lost about $700 million in its largest fund due to extreme short positions.[9] In November 2018, OptionSellers.com went bankrupt because it suffered losses of more than 150 million USD in trading commodity options. The company's investment strategy was shorting strangles, which resembles the payoff profile of a delta-hedged short option position. Of course, OptInts are not comparable with funds. While the latter relies on investment profits, OptInts' main income is spread income. Nevertheless, OptInts must take some risks and follow dynamic strategies (Baird, 1992). OptInts have intense monitoring and risk management systems to manage the risks associated with making the market. (Bergomi, 2016).

The risk management of options has also been integrated into the regulatory frameworks in the USA and Europe. Title 12 Part 44 of the Code of Federal Regulations (CFR) establishes the

[9] See the news release by the SEC: `https://www.sec.gov/news/press-release/2020-21`. More detailed reports by the SEC and CFTC mention that the fund relied on selling call options and that the risk management was inadequate.

legal framework for the USA. The framework has been amended in response to the Dodd-Frank Act and is integrated into the Bank Holding Company Act. Among other information, the rules require reporting a comprehensive P&L attribution that contains the following information: the actual daily P&L, the P&L from existing and new positions, the P&L attributed to changes in risk factors, and the market values of the current long and short positions. Most importantly, the regulator requires to report risk limits that define the amount of risk a trading desk takes. The limits can be quantified with the 1% Value-at-Risk, positional limits, or stress scenarios.

The procedures are often only vaguely defined and refer to the internal risk management techniques of the intermediary. More details are provided in the Basel III framework (Bank for International Settlements, 2019). The framework considers only three sources of risks: directional risk (Δ), convexity risk (Γ), and volatility risk (ν). For the quantification of convexity, the Basel committee requires using stress scenarios (MAR 10.16). The approach is based on stressing the current option positions by assuming a $\pm 8\%$ move in the underlying that is accompanied by a $\pm 25\%$ shift in implied volatility (MAR 40.83). The capital requirements are determined based on the highest loss along the grid of scenarios. The framework does not explicitly define calculation methods for the sensitivities. However, MAR 21.28 mentions that the volatility sensitivity of equity options can be calculated using a log-normal assumption, which is in line with Black and Scholes (1973). The maximum losses of the scenarios are taken to determine the capital requirements. Interestingly, the capital requirements for Γ are only for net short positions (MAR 40.80).

C. Quantifying Unhedgeable Risks via Stress Scenario

I use stress scenarios to determine the OptInts' P&L in case of a shock. The approach recognizes that OptInts are mostly concerned about actual capital losses and unifies most unhedgeable risks into one quantity. More specifically, I define the intermediary risk as the potential loss for different moves in the underlying and implied volatility. The delta-neutral P&L for each scenario i for the next trading day is simply given by

$$E\left[\text{P\&L}_{t+1}^{i}\right] = \sum_{j} Pos_{j,t} \times \left(\underbrace{(O_{t+1}^{j,i}(S^i, K, IV^i, \tau - \delta t) - O_t^j)}_{\text{Scenario Shock}} - \underbrace{\Delta_t^j (S^i - S_t)}_{\text{Hedge}} \right). \tag{14}$$

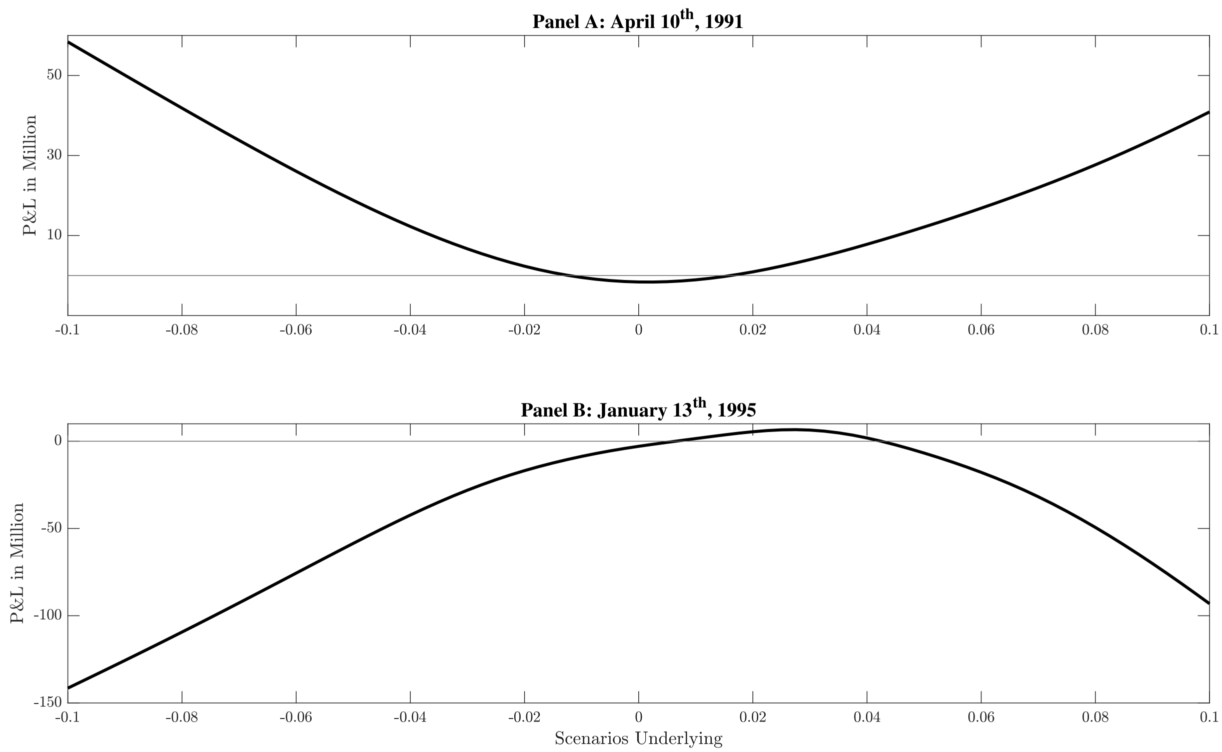
Hence, I use the current positions in option j, shock the option value O_t^j by assuming scenarios for the underlying S^i and the implied volatility IV^i, and determine the delta-hedge P&L for each scenario. To implement the method, I require a grid of scenarios and a pricing model to determine the option values along the grid.

I use a range of ±10% for changes in the underlying, slightly extending the Basel III suggestions. The reason is that nine daily returns have been outside the suggested range of ±8% over the last 15 years. I use three different implied volatility scenarios. First, I follow a simple but common practice of trading desks and use the 'sticky-strike' rule. The rule assumes that implied volatilities will be unaffected by changes in the underlying, preserving the current volatility surface (Daglish, Hull, and Suo, 2007). Second, I assume that a change in the underlying shifts the implied volatility by -3.65, matching the unconditional relationship between S&P 500 and VIX returns. The Basel Committee suggests a similar scenario. Conveniently, the magnitude of the volatility shifts will depend on the direction in which the underlying moves. Negative underlying changes will increase the implied volatility and vice versa. Lastly, I extend the approach and use a 120-day rolling window regression coefficient between returns in S&P 500 and VIX. This accounts for the time-varying relationship and the current state of the economy. The last step to determining the scenario P&L is the choice of the pricing model. For the underlying purpose, it is essential to note that the model does not need to be designed to create a single reference distribution that matches the observable volatility surface. Instead, the model takes current surface characteristics (volatility skew, term structure) and reprices options given a change in the underlying or implied volatility. Hence, the model does not need to be as complex as modern option pricing models. I choose the simple BSM model. Figure G.3 confirms that the model is a reasonable choice. The mean pricing error over the whole sample period is 1%. Hence, only a tiny fraction of the observable option prices cannot be explained using the BSM model and the current volatility surface.

Figure 4 plots the scenario P&L for two random days. As is evident, the scenario P&L is just a weighted average of individual options delta-hedged P&L. Panel A shows the scenario P&L for April 10$^{\text{th}}$, 1991, for which the OptInts' position was net long. OptInts make a slight loss when the market does not move too much (loss in time value) but profit in case volatility is high. The portfolio is not fully balanced across strike prices, resulting in an asymmetric P&L. The potential gain is slightly more when markets go down, as indicated by the steeper slope for negative underlying changes. Panel B shows the P&L for January 13$^{\text{th}}$, 1995. The positions are net short such that OptInts make a loss when volatility is high in exchange for the gain in time decay. Although option markets are considerably smaller, the losses amount to over $100 million for some scenarios. The profit is approx. $10 million when the S&P 500 increases by 2-3%.

The figure also indicates that losses are potentially unlimited, exposing OptInts to the risk of bankruptcy and affecting the whole system. Brunnermeier and Sannikov (2014) show that option markets may have led to more frequent crises. Intermediaries (and other participants) could attain risky positions with high leverage but did not have to maintain large capital buffers. I visualize the evolution of the scenario P&L for the two most extreme scenarios in Figure 5.

Figure 4. Market Maker Scenario P&L



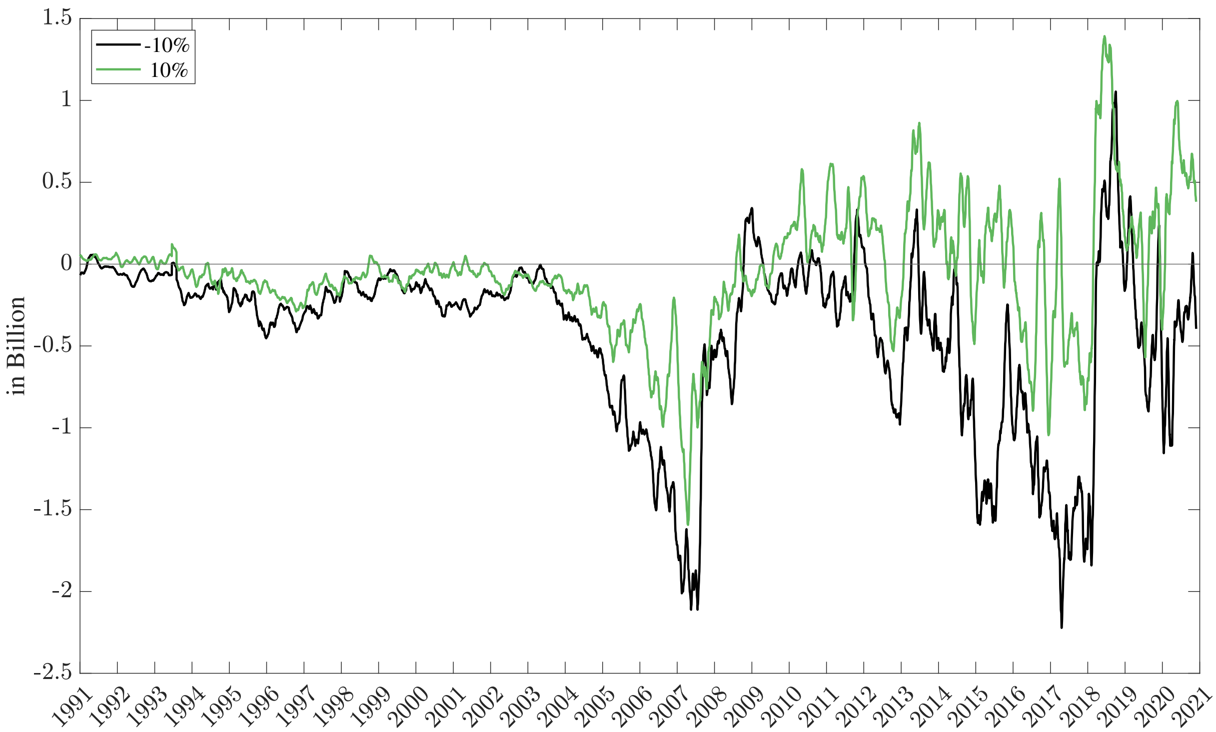
The figure exemplifies Scenario P&Ls for two randomly selected dates. The P&L is determined via Equation (14) and by assuming changes in the underlying along the grid of ±10% and shifts in implied volatilities of 3.65 times the change in the underlying.

Upward shocks often result in capital losses for intermediaries, but downward shocks are always more painful. The magnitude of losses is extreme, and intermediaries are exposed to daily losses of more than $2 billion.

D. Option Intermediary Risk Factor - Scenario P&L Slope

Regulators require to use the maximum loss based on a two-shock stress scenario to determine the regulatory capital. While this makes sense from a regulator's point of view, I suggest using the scenario P&L's negative slope. Slopes have the appealing property that unlimited losses are taken into account. A slope describes the direction of a function at a particular point. If the slope is negative, OptInts will make losses relative to the previous scenario. Given that the usual shape of a scenario P&L is convex or concave, it indicates the functional form. Additionally, the slope quantifies the steepness of the P&L at a given point. Larger absolute values indicate that the rate of change of the P&L in the stress scenario is high. Therefore, the slope provides an

Figure 5. Time-Series of Extreme Scenario P&L



The figure shows 30-day moving averages of two scenario P&Ls. The scenarios are $\pm 10\%$ changes in the underlying and fixed shifts in implied volatilities of 3.65 times the change in the underlying.

intuitive description of the exposure to very large adverse shocks. A potential disadvantage is that a slope does not consider the current level of the P&L. That is, the factor can be downward sloping, although OptInts currently make a profit and vice versa. Below, I show that this is not of concern.

To determine the slope, I need a meaningful reference point. I quantify how risky OptInts' P&L is when the stock market crashes. There is abundant evidence that economic turmoils and downside states are of primary interest to the economy, policymakers, and asset markets.[10] Such episodes are also of first-order importance for OptInts because they present a shock to intermediary constraints and have higher loss potential. I follow the Basel III framework and use the scenario with the highest loss. It turns out that the downward shock of -10% is always the minimum of Equation (14).

The size of option markets has increased drastically over the last few years. I standardize the slope factor by the absolute number of positions to counteract the severe time trend. With the

[10]See, among many others, Gabaix (2012) for a model with rare disaster, Kelly and Jiang (2014) for stock returns, and Andersen, Fusari, and Todorov (2015b) for option markets.

standardize, I account for the growth in the market and the fact that more intermediaries make the option market.[11] Consequently, the negative slope factor is given by

$$OIR_t = \frac{\text{P\&L}_{t+1}^{-10\%} - \text{P\&L}_{t+1}^{-10\%+z}}{|\sum_j Pos_{j,t}|}, \tag{15}$$

where $z = 0.01\%$. Hence, the option intermediary risk OIR is the difference between two adjacent scenario P&Ls standardized by the absolute number of positions. The factor can be regarded as the additional profit or loss per unit of option contract held in case a crash occurs.

Table I provides summary statistics for the unstandardized factor OIR^{raw}, OIR, and the scenario P&L the different implied volatility scenarios. All factors are scaled. The mean and median (Q50) factors are negative for all specifications. For example, the mean OIR^{raw} the fixed IV shift scenario equals an additional \$8.2 million. The numbers shift downward for the sticky-strike scenario and upward for the dynamic-IV scenario. The factors show considerable time-series variation. The estimated standard deviation is almost one and a half the mean. The difference between the mean and median indicates a severe left skew. The minimum OIR^{raw} for the fixed IV shift scenario is more than \$62 million. Sometimes, OptInts would profit if the underlying crashed, as indicated by the maximum values. Panel B shows that the summary statics are similar for OIR. Two notable exceptions are the significant decrease in standard deviation and skewness. Panel C highlights that the summary statistics for $\text{P\&L}_{t+1}^{-10\%}$ align well with the slope factors. The mean loss for the fixed IV shift scenario is \$430 million, and the maximum potential loss equals \$3.6 billion. As documented by Gârleanu, Pedersen, and Poteshman (2009), daily realized losses of more than \$100 million occurred between 1996 and 2001, although the market was smaller than nowadays. The signs of the slope factor and the P&L are equal for more than 95% of the sample. The correlation is about 93%. While cases with opposite signs do not often occur, they are informative for the OptInts because they indicate whether losses or profits are limited. A profit will turn into a loss when markets crash even more severe in $\approx 4\%$ of observed days, and a loss turns into a profit in $\approx 1\%$ of observed days.

Figure 6 plots one-week moving averages for OIR for all three implied volatility scenarios. Indeed, OIR is primarily negative, implying that OptInts almost always have a negative P&L when markets crash. This suggests that OptInts are averse to big market crashes. Interestingly, the plot reveals that OptInts often build up more exposure against stock market jumps before crises but are relatively unaffected once the market actually crashes. This was the case before the financial crisis in 2008 and the Corona crisis in 2020. The highest maximum loss for the dynamic scenario (and for all three raw measures) occurred at the beginning of 2018, right before

[11] There is no readily available information about the evolution of the number of OptInts. However, Gârleanu, Pedersen, and Poteshman (2009) report about 100 OptInts, while the OCC currently lists 140.

Table I. Summary Statistics

	Mean	SD	Min	0.1	0.25	0.50	0.75	0.90	Max
				Panel A: OIR^{raw}					
Sticky IV	-53.68	79.05	-457.25	-168.25	-78.20	-25.44	-8.67	6.14	188.35
Fixed IV	-82.29	106.39	-625.54	-250.80	-120.72	-37.68	-15.80	-2.16	143.92
Dynamic IV	-118.57	176.56	-1261.09	-391.03	-158.69	-38.76	-15.54	-2.05	118.23
				Panel B: OIR					
Sticky IV	-0.027	0.026	-0.130	-0.063	-0.042	-0.022	-0.007	0.002	0.038
Fixed IV	-0.039	0.033	-0.158	-0.087	-0.061	-0.032	-0.013	-0.001	0.029
Dynamic IV	-0.049	0.046	-0.245	-0.121	-0.073	-0.036	-0.015	-0.001	0.025
				Panel C: P&L					
Sticky IV	-19.53	39.77	-226.38	-68.62	-32.08	-12.29	-2.76	12.95	219.12
Fixed IV	-42.78	56.29	-366.84	-129.04	-63.03	-23.63	-9.53	1.87	215.36
Dynamic IV	-70.38	104.30	-1018.87	-231.58	-94.13	-26.11	-10.05	0.05	214.87

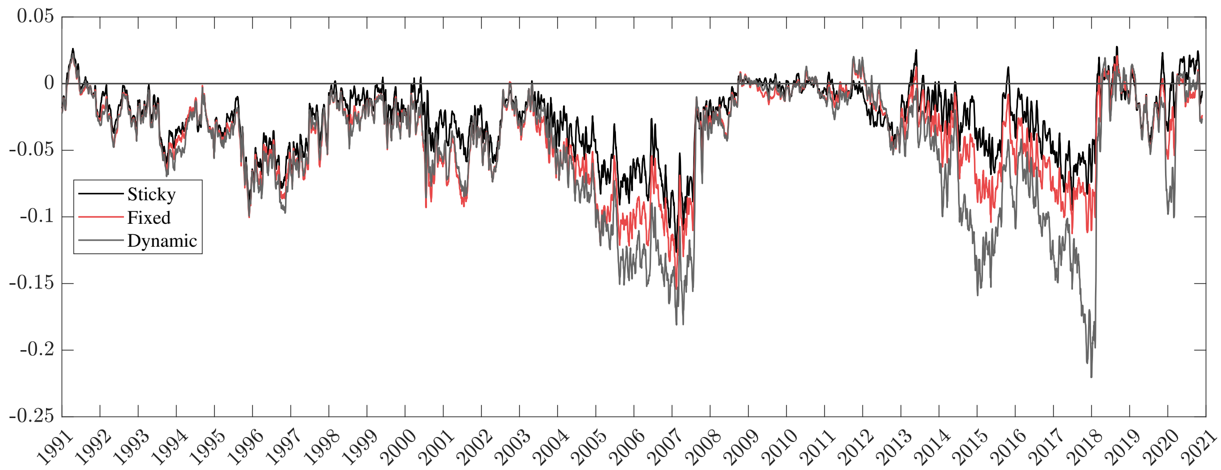
This table provides summary statistics for unstandardized option intermediary risk OIR^{raw} defined as ($\text{P\&L}_{t+1}^{-10\%} - \text{P\&L}_{t+1}^{-9.99\%}$) and scaled by 10^{-5} (in Panel A), OIR defined in Equation (15) and scaled by 10^{-2} (in Panel B), and the scenario $\text{P\&L}_{t+1}^{-10\%}$ defined in Equation (14) and scaled by 10^{-7} (in Panel C). The IV scenarios are sticky-strike (Sticky IV), fixed IV shift of -3.65 multiplied with the stock price change of 10% (Fixed IV), and a dynamic shift estimated from a 120-day rolling window of VIX returns on S&P 500 returns (Dynamic IV).

a volatility fund's bust amid a sharp volatility increase ('Volmageddon'). I observe no material difference in the time-series dynamics across OIR across the different scenarios. Regardless of the construction, the dynamics are very similar. The most notable exception is the severe decrease in the dynamic implied volatility time series at the end of 2017. Pairwise correlations are 95% between the sticky and fixed scenario, 85% for sticky and dynamic, and 94% for fixed and dynamic. The raw versions have slightly higher correlations. Throughout the paper, I will focus on the fixed IV shift scenario. This is in line with the Basel Committee's suggestion. Additionally, the factor has the highest correlation with the other two factors. I show below that the results do not depend on the exact specification of the intermediary risk factor (P&L level or slope), the reference point, or the implied volatility scenario.

IV. RISK AND LIQUIDITY PREMIA

In this section, I study the information content of intermediaries' positional risk. I start by discussing my empirical identification and my instrumental variable approach. I show that past order imbalances, option expiry days, and intermediary constraint as instruments can be instrumented for my option intermediary risk factor. The next two section relate OptInts to liquidity premia calculated from realized S&P 500 option transactions and delta-hedged option returns.

Figure 6. Time-Series of OIR for different IV Scenarios



The figure shows the OIR from Equation (15) from 1991 to 2021. The factor is determined for the sticky-strike IV scenario (black line), fixed IV shift scenario (blue line), and dynamic IV shift scenario (gray line). The stock price scenario is set at -10%.

A. Identification: Lags and Instrumental Variable

At any given point in time, the information in my risk factor OIR consist of three sources: the aggregate net positions Pos_t, the current option surface, and the scenarios that approximate tomorrow's option surface in case of a shock. I use this measure to establish a link between the risk of intermediating the option market and its premia. The dependent variables are delta-hedged option returns and effective option spreads. Delta-hedged option returns load on risks that are not spanned by trading the underlying, such as variance and jump risk (Bakshi and Kapadia, 2003). Spreads are a proxy for the compensation for providing liquidity (Christoffersen, Goyenko, Jacobs, and Karoui, 2018). The exact definition of both variables is given in the respective section.

Fixing the dependent and independent on the same timeline will likely result in biased estimates. Changes in Pos_t and option prices are endogenous due to simultaneity and concurrent factors such as news about fundamentals affect both. Are prices increasing because investors buy or vice versa? Are spreads decreasing because investors sell? I will address this issue in two ways. First, I use lagged dependent variables because it is unlikely that demand in t will cause further price pressure the next day. Among others, Muravyev (2016) analyzes the effect of demand on option prices and uses an identification based on instantaneous price changes. Hendershott and Seasholes (2007) use lags to identify the effect of stock market makers' positions on future price changes. The underlying assumption is that the simultaneity between quantities in prices is instantaneous. A simple test is to check for time dependencies in the dependent variables. I find that the average 1-day autocorrelation across the panel of delta-hedged option returns is 0.02 and statistically insignificant. The model I run for the lag identification is

$$Y_{i,t} = \alpha + \beta OIR_{t-1} + \gamma X + \phi_t + \varphi_i + \epsilon_{i,t}, \tag{16}$$

where $Y_{i,t}$ are option returns (spreads) for different portfolios, X are a set of controls, ϕ_t are year-month fixed effects, and φ_i are entity fixed effects. Hence, the model controls for heterogeneity. Standard errors are double-clustered at entity and time levels.

Second, I apply an instrumental variable approach by instrumenting past order imbalances, option expiries, and intermediary constraints. Order imbalances are the residual between buy and sell transactions. More buy activity results in a positive order imbalance which intermediaries must absorb. The model of Chordia and Subrahmanyam (2004) shows that past order imbalances predict future changes in the aggregate position of intermediaries. Past order imbalances predict variation in Pos_t and, thus, in OIR. Similarly, expired options create exogenous variation as the expired options are not further in the intermediary's positions. Muravyev (2016) uses both variables to instrument current order imbalances. I use rapid increases in the leverage ratio measure of He, Kelly, and Manela (2017) as a third instrument. Increases in the leverage ratio indicate tighter intermediary constraints. This should lead to variation in OIR through changing option prices as OptInts are less willing to supply options. The leverage ratio is determined from all NY Fed primary dealers' average market equity and book debt. The largest primary dealers are also OptInts.[12] I use a dummy that equals one if changes in the leverage ratio are larger than 2.5% to proxy for tighter constraints of intermediaries. Because concurrent factors might drive changes in the leverage ratio and option prices, I lag the dummy variable by one week.

The variables must satisfy three conditions to be valid instruments. First, the instruments cannot be correlated with the dependent variable other than through the variation in intermediaries' positions. I apply the Wu-Hausman test, and find no evidence for endogeneity between the instruments and the outcome variables. I conclude that the condition is satisfied. Second, the instruments should not correlate to an omitted variable that determines option returns. It is unclear whether past order imbalances predict only future changes in intermediaries' positions or contain information about future volatility (informed trading). Muravyev (2016) addresses this concern and shows that past order imbalances primarily drive future positions rather than future volatility. Third, the instrument must be strongly correlated with the endogenous explanatory variable. I apply the following first-stage regression to test this condition:

$$OIR_t = \alpha_0 + \alpha_1 OI_{t-7} + \alpha_2 OI_{t-6} + \alpha_3 OI_{t-5} + \alpha_4 OI_{t-4} + \alpha_5 OI_{t-3} + \alpha_6 Exp_t + \alpha_7 Constraint_{t-7} + \epsilon_t. \tag{17}$$

OI_{t-n} is the residual between buying and selling orders, Exp_t is a dummy that equals one

[12] See the list of NY Fed Primary Dealers and the Option Clearing Corporation member directory.

when t is an option expiry day, and $Constraint_{t-7}$ is a lagged dummy that indicates if NY fed primary dealer's leverage increased by more than 2.5%. I do not use the most recent order imbalances and constraint dummy because both could be driven by the same variable that drives option qua. The results are reported in Table II. The table shows that all coefficients in the first stage are highly significant and correlated to OIR. The results highlight that each instrument adds isolated explanatory power, confirming that each instrument is correlated to OIR. Hence, the instruments are a good choice. The coefficients for OI are negative because positive order imbalance results in new short positions for OptInts, thereby decreasing OIR (higher potential losses). The coefficient for the expiry dummy is positive. On average, the expired options reduce the short positions of OptInts and the overall loss potential. The lagged dummy for intermediary constraints is positive, indicating that higher constraints are associated with lower values of intermediary risk. The most likely channel is that intermediaries have to reduce risky short positions when they are more constrained.

Table II. First Stage Regression of Instrumental Approach

	OIR_t	OIR_t	OIR_t
OI_{t-7}	-0.30^{***}	-0.32^{***}	-0.28^{***}
	(0.05)	(0.05)	(0.05)
OI_{t-6}	-0.29^{***}	-0.32^{***}	-0.29^{***}
	(0.05)	(0.05)	(0.05)
OI_{t-5}	-0.30^{***}	-0.34^{***}	-0.32^{***}
	(0.05)	(0.05)	(0.05)
OI_{t-4}	-0.32^{***}	-0.31^{***}	-0.31^{***}
	(0.05)	(0.05)	(0.05)
OI_{t-3}	-0.32^{***}	-0.32^{***}	-0.32^{***}
	(0.05)	(0.05)	(0.05)
Exp_t		1.30^{***}	1.30^{***}
		(0.30)	(0.29)
$Constraint_{t-7}$			2.37^{***}
			(0.33)
adj. R^2	8.19	10.46	12.36
F-Stat	75	81	84

The table reports results of the first stage regression from Equation (17). OI is the order imbalance defined as the residual between buy and sell transactions. Exp is a dummy that indicates option expiry days. $Constraint$ is a dummy that equals one when the leverage ratio of financial intermediaries increased by more than 2.5%. Standard errors (in parentheses) are computed based on the method of Newey and West (1987) with 30 lags. ***, **, and * denote significance at 1%, 5%, and 10%. The sample period is from Jan 2004 until Dec 2020. The dependent variables are standardized by their sample standard deviation and multiplied by 100. The regression constant is not reported.

B. Impact on Liquidity Premia

I first focus on liquidity costs in option markets. OptInts' compensation for providing immediacy is the so-called spread which is the difference between the transaction and fair price. For every buy (sell), OptInts offers a price above (below) the option's fair price. Higher spreads indicate less market liquidity because OptInts require more compensation to provide immediacy. I construct liquidity measures from traded S&P 500 options at the CBOE. The data reports relevant trade conditions, such as the exact time stamp, trade size, price, best bid/ask, and trade IV. The data is available from 2004 until the end of 2020. The liquidity measure I focus on is along the lines of Christoffersen, Goyenko, Jacobs, and Karoui (2018), who use the effective relative dollar spread of an option trade. However, I deviate from the proposed measure by using implied volatility spreads. Implied volatilities are more comparable than dollar values. A \$0.01 spread for a \$0.05 option is relatively higher than a \$5 spread for a \$50 option. Implied volatility spreads ensure comparability across different contracts. The measure is defined as

$$ES^k = \frac{2|IV_k^P - IV_k^M|}{IV_k^M}, \tag{18}$$

where IV_k^P is the implied volatility of option trade k calculated with the trade price P, and IV_k^M is the implied volatility of the same option calculated with the mid-price M. I construct five different moneyness buckets b for calls and puts separately.[13] Each trade's contribution to the bucket spread is weighted by its volume. The median absolute effective IV spread across the panel is 0.018. Table III reports the results of the panel regressions.

Intermediary risk and option liquidity are strongly related. The coefficient is consistently negative and statistically meaningful. Intermediaries' positional risk well captures the variation of spreads over time. The relationship is more important (and less spurious) than the VIX and stock market illiquidity. Funding illiquidity is also an important contributor to explaining option market liquidity. Adding the previous days' effective relative spreads and the traded option volume as controls reduces the magnitude of OIR, but the statistical importance is unaffected. In contrast, the VIX and stock market illiquidity become insignificant, and the coefficient of FI reduces by almost 80%. Using the instrument for intermediary risk shows that the effect is indeed negative and causal. On average, the relationship between intermediary risk and liquidity cost is negative, suggesting that OptInts require higher compensation for providing liquidity on days on which the P&L is risky. The findings are in line with Ho and Stoll (1981), who show theoretically that (stock) market makers' positions are negatively correlated with spreads. The intuition is that spreads are adjusted so that positions are less likely to become even more

[13]The five buckets are deep out-of-the-money $b < 0.90$, out-of-the-money $0.90 \leq b < 0.975$, at-the-money $0.975 \leq b < 1.025$, in-the-money $1.025 \leq b < 1.10$, and deep-in-the-money $b > 1.10$.

Table III. Liquidity Compensation

	$ES_{i,t}$	$ES_{i,t}$	$ES_{i,t}$	$ES_{i,t}$
OIR	-0.282^{***}	-0.245^{***}	-0.108^{***}	-0.019^{***}
	(0.065)	(0.064)	(0.034)	(0.005)
VIX_t		-0.278^{***}	-0.019	-0.091^{**}
		(0.105)	(0.037)	(0.044)
SI_t		0.098^{**}	0.002	-0.007
		(0.046)	(0.022)	(0.021)
FI_t		0.739^{***}	0.162^{***}	0.205^{***}
		(0.116)	(0.051)	(0.061)
$ES_{i,t-1}$			0.457^{***}	0.467^{***}
			(0.053)	(0.055)
$ES_{i,t-2}$			0.403^{***}	0.412^{***}
			(0.047)	(0.050)
$Volume_t$			-0.259^{***}	-0.260^{***}
			(0.085)	(0.086)
adj. R^2	9.10	14.68	31.04	30.81
adj. R^2 w/o	2.09	7.59	17.31	16.81
Identification	Lag	Lag	Lag	Inst
Entity FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
N	41,000	41,000	41,000	41,000

The table reports results of panel regressions of IV spreads from Equation (18) on the option intermediary risk factor OIR and several controls. VIX is the CBOE's volatility index, SI is the Amihud (2002) market illiquidity measure, FI is funding illiquidity of Hu, Pan, and Wang (2013), ES_{t-n} are lagged spreads, and $Volume$ is the traded option volume. Lag indicates that $m^{\text{P\&L}}$ is lagged by one day. Inst is the instrument from Equation (17). R^2 w/o is the explained variation without fixed effects. Standard errors (in parentheses) are clustered by entity and month. ***, **, and * denote significance at 1%, 5%, and 10%. The sample is from Jan 2004 until Dec 2020. Independent variables are standardized and multiplied by 100. The regression constant is not reported.

negative (or positive). Similar evidence for a small sample of equity options is provided by Ho and Macris (1984).

C. Impact on Risk Premia

OptInts across assets strongly desire to exit markets with no position (Hendershott and Seasholes, 2007). One way to control the positions is to adjust spreads around the fair price determined by the fundamental forces affecting supply and demand. A fundamental force may be the risk of OptInts which are usually net suppliers. OptInts may account for the risk by adjusting both spreads and prices. Because option prices are not comparable over time, I use delta-hedged option returns as a price measure. I follow the insights of Muravyev and Ni (2020) and Goyenko and Zhang (2020), who show that overnight option returns are strikingly different from intraday returns. I construct three sets of option returns for different investment horizons: end-of-day returns (4:00 pm to 4:00 pm), overnight returns (4:00 pm to 9:40 am), and intraday returns (9:40 am to 4:00 pm). I use 4:00 pm because asynchronous and stale underlying prices are otherwise an issue.[14] Similarly, option quotes are relatively stale at the beginning of trading at 9:30 am (Muravyev and Ni, 2020). I define the respective delta-hedged return of option O as

$$r_{t+1} = \frac{O_{t+1} - O_t - \Delta_t(F_{t+1} - F_t)}{O_t}, \tag{19}$$

where F denotes the underlying's future price and Δ is the option's delta. The option returns are constructed for close-to-expiry options (less than 13 days TTM), short-maturity options (30 days), medium-term options (150 days), and long-term options (>150 days). Moreover, I categorize option returns by option type and five moneyness buckets defined in Section B. Results of regressions are given in Table IV.

Tomorrow's option returns strongly depend on potential capital losses. OIR is statistically highly significant in predicting end-of-day, intraday, and overnight option returns. A one standard deviation increase is associated with returns of -1.78% day-to-day, -1.40% intraday, and -0.53% overnight using the lag identification. The coefficient's magnitude reduces for the instrumental variable approach, but all coefficients are still economically large and statistically meaningful. The causal effect is an option return of more than -0.50%. OIR impacts returns through a different channel than volatility-risk weighted positions Vega. An increase in Vega (primarily negative) predicts a positive option return, while an increase in OIR predicts a negative option return. The market-wide order imbalance across all option transactions is also statistically highly significant. This is in line with previous research on option returns and demand (Muravyev, 2016; Muravyev and Ni, 2020). However, although the variable is not lagged, the magnitude of OIR is four to six times larger (up to three for the instrumental variable approach). The economic significance of aggregate positional risk exceeds order imbalances by far. However, the coefficient of OI is likely biased due to endogeneity. In unreported results, I find that the predictability lasts up to 3 days, suggesting that intermediaries adjust prices gradually when positional risk is high. The results highlight that OptInts want to induce public sell orders by charging higher prices and spreads.

[14] Equity markets usually stop trading at 4:00 pm while options stop trading at 4:15 pm. The consequences and importance are discussed in the Appendix of Battalio and Schultz (2006). Note that the results do not depend on the exact time.

Table IV. Option Returns

	$r_t^{end-of-day}$		$r_t^{intraday}$		$r_t^{overnight}$	
	(1)	(2)	(3)	(4)	(5)	(6)
OIR	-1.777^{***}	-0.511^{**}	-1.395^{***}	-0.277^{***}	-0.525^{**}	-0.275^{***}
	(0.549)	(0.215)	(0.423)	(0.102)	(0.209)	(0.099)
VIX	-1.744^{**}	-1.668^{**}	-0.209	-0.216	-1.076^{***}	-0.990^{***}
	(0.821)	(0.785)	(0.550)	(0.507)	(0.411)	(0.376)
$Vega$	1.152^{**}	0.642	0.796^{**}	0.372	0.690^{**}	0.594^{**}
	(0.582)	(0.508)	(0.383)	(0.308)	(0.271)	(0.241)
VRP	-0.120	-0.174	-0.053	-0.048	-0.027	-0.077
	(0.451)	(0.459)	(0.314)	(0.304)	(0.115)	(0.125)
$Skew$	0.010	0.006	-0.042	-0.090	0.163	0.180
	(0.601)	(0.585)	(0.371)	(0.371)	(0.232)	(0.228)
r_t^{SP500}	-3.253^{**}	-3.266^{**}	-2.421^{***}	-2.421^{***}	-2.211^{***}	-2.219^{***}
	(1.296)	(1.298)	(0.859)	(0.859)	(0.830)	(0.832)
adj. R^2	8.99	8.97	4.72	4.68	14.79	14.82
adj. R^2 w/o	3.48	3.46	3.36	3.34	4.84	4.84
Identification	Lag	Inst	Lag	Inst	Lag	Inst
Entity FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
N	132,259	132,259	132,259	132,259	132,259	132,259

The table reports results of panel regression of end-of-day, intraday, and overnight option returns on option intermediary risk OIR and several other predictors/controls. VIX is the CBOE's volatility index, $Vega$ are the volatility risk-weighted positions, VRP is the variance risk premium, and $Skew$ is the risk-neutral skewness, r^{SP500} is the return of the S&P 500 calculated over the same period as the option returns, and OI is the market-wide order imbalance defined as the sum over all buy and sell volumes. Identification Lag indicates that $m^{\text{P\&L}}$ and the other variables without a time-index are lagged by one day. Identification Inst use the instrument from Equation (17). R^2 w/o is the explained variation without fixed effects. Standard errors (in parentheses) are clustered by entity and time. ***, **, and * denote significance at 1%, 5%, and 10%. The sample period is from Jan 2004 until Sep 2020. All independent variables are standardized and multiplied by 100. The top and bottom 0.5% outliers of all sets of returns are removed. The regression constant is not reported.

D. Robustness Checks

I report various robustness checks for end-of-day option returns in Table G.3 and effective spreads Table G.4. I use the lag and the instrumental variable identification for all robustness checks. I note that the coefficient estimated via instrumental variables is almost unchanged throughout all option return robustness checks.

Reference Point: The first two columns report results when OIR is determined at the scenario -5% instead of -10%. The coefficient for the lag identification decreases slightly so that a one standard-deviation shock in intermediary risk OIR is associated with a delta-hedged option return of 1.20% the next day (instead of 1.77% in the main results). However, the instrumental approach reports a slightly more negative coefficient – the coefficient decreases from -0.51 to -0.53. The results for effective spreads improve significantly for both identification approaches. For instance, the instrumental variable coefficient decreases from -0.02 to -0.05. I also test further reference points and find that every scenario $< -2\%$ produces the same results. Interestingly, the coefficient's sign changes for positive scenarios.

P&L Level: The next two columns report results when I use the level $E\left[\text{P\&L}^i_{t+1}\right]$ from Equation (14) determined at scenario -10% instead of the slope factor. The results are unchanged.

Implied Volatility Scenario: The next robustness shows that the results are robust to the implied volatility scenario. I determine OIR using the dynamic IV scenario. I estimate the relation between underlying and implied volatility changes via a 120-day rolling window regression of S&P 500 returns on VIX returns. The results slightly improve.

Volatility States: I remove all days with high volatility from the sample to check whether the main results only hold during volatile periods. High volatility days are defined as $VIX > 20\%$. The results for option returns do not change. The instrumented coefficient decreases from -0.51 to -0.56. For spreads, the coefficients within the lag identification are similar to the main results. The coefficient within the instrumental identification is insignificant. It appears that intermediaries predominantly increase spreads in times of high volatility.

Realized Profit&Loss: A natural concern is whether results are robust to realized profits or losses of intermediaries. I construct a time series of realized profits or losses, assuming that intermediaries delta-hedge their positions at the end of each trading day. The coefficient of the realized P&L is large and statistically highly significant. A one-standard deviation decrease is associated with returns of about 2.5% but is probably inflated due to endogeneity. Nevertheless, the inclusion does not affect my intermediary risk factor. The coefficient in the instrumental approach is virtually unchanged. With regard to realized spreads, the realized P&L has no significant explanatory power.

V. TIME-VARIATION IN OPTION INTERMEDIARY RISK

In this section, I show that option intermediary risk varies with the health and constraints of financial institutions, the economic outlook, volatility. Then, I analyze how OptInts manage OIR and document asymmetrically adjusted option quotes and spreads that induce public sell orders.

A. Intermediary Health and Volatility

My measure is constructed to capture the risk of intermediating in the option market. Many OptInts are large financial institutions also active in other markets. Brunnermeier and Pedersen (2009) show that the funding of intermediaries and market liquidity are mutually reinforcing, and traders' ability to provide liquidity depends on funding availability and vice versa. I compare my risk measure to a variety of factors. The first is the funding illiquidity measure of Hu, Pan, and Wang (2013), which exploits deviations of observable bond yields to model-implied bond yields. The second measure is the value-weighted leverage ratio of NY Fed Primary Dealers from He, Kelly, and Manela (2017). Primary dealers are the largest financial institutions that are trading counterparties for the Fed in its implementation of monetary policy. The largest primary dealers are also OptInts. The leverage ratio is defined as total assets divided by equity and is a well-known proxy for the health of the intermediary sector (He and Krishnamurthy, 2013). The third measure is the difference between the 3-month LIBOR and Treasury rate (TED spread), a proxy for the credit risk of financial intermediaries. The fourth measure I use is the difference between the yield of the AAA and BBB BofA Corporate Index. The yield spread proxies the aggregate default risk and is a strong predictor of economic activity (Culp, Nozawa, and Veronesi, 2018). All measures are publicly available at a daily frequency. Table V depicts the relationship to my option intermediary risk factor.

Table V. Relation to Intermediary Health

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
α	-1.56^{***}	-2.10^{***}	-1.32^{***}	-2.08^{***}	-2.37^{***}	-2.36^{***}	-2.40^{***}
	(0.10)	(0.14)	(0.08)	(0.14)	(0.17)	(0.17)	(0.17)
FI_t	0.28^{***}				-0.08		-0.11
	(0.05)				(0.06)		(0.07)
LR_t		0.49^{***}			0.26^{***}	0.26^{***}	0.25^{***}
		(0.06)			(0.05)	(0.06)	(0.05)
TED_t			0.11^{***}			-0.02	0.04
			(0.04)			(0.05)	(0.06)
DS_t				0.47^{***}	0.40^{***}	0.35^{***}	0.41^{***}
				(0.06)	(0.08)	(0.07)	(0.08)
adj. R^2	7.8	23.7	1.2	21.8	30.9	30.6	31.1
N	7467	5430	7337	6018	5386	5291	5287

The table reports results of daily OLS regression of option intermediary risk OIR on measures of financial stability. FI is the funding illiquidity measure of Hu, Pan, and Wang (2013), LR is the value-weighted debt-to-equity ratio of NY Fed Primary Dealers of He, Kelly, and Manela (2017), TED is the difference between the 3-month LIBOR rate and treasury yield, and DS is the difference between the AAA and BBB BofA Corporate Bond Index. Standard errors (in parentheses) are computed based on the method of Newey and West (1987) with 30 lags. ***, **, and * denote significance at 1%, 5%, and 10%. The sample period is from Jan 1991 until Dec 2020, except for all regressions including LR (May 1999) and DS (Jan 1997). All variables are standardized by their sample standard deviation.

The first regression shows that the option intermediary risk positively relates to the illiquidity proxy. More funding illiquidity, indicating a lack of arbitrage capital, is associated with less severe losses / a profit of OptInts in case of an extreme adverse shock. Potential losses from the option position are also positively related to the leverage ratio of NY Fed Primary Dealers. A one standard deviation increase in LR translates into a 0.49 standard deviations increase in OIR. A high leverage ratio suggests that operational risk is high and intermediary constraints are likely to be binding. The results indicate that this translates to the option trading desks because OptInts reduce potential losses. The large explained variation suggests a strong relationship between both variables. Option intermediation risk is also related to the TED spread, but the R^2 is low compared to the other variables. The fourth regression shows that OIR is strongly associated with the default spread. A higher spread, indicating a bad economic outlook, is associated with lower worst-case losses for OptInts. The regressions with multiple independent variables highlight that the leverage ratio and the yield spread are essential for OIR. Both measures are statistically significant throughout all regressions and create a sizable R^2 of more than 30%. Both variables provide unique information. OptInts reduce their exposure when financial constraints are binding and when the economic outlook is pessimistic. The risk of option intermediation depends not only on option-specific factors but also on the overall health of the financial sector and the economic outlook. Both translate directly to the option market. Tighter constraints are associated with lower potential losses resulting from OptInts' reducing short positions. Traders become averse to taking positions.

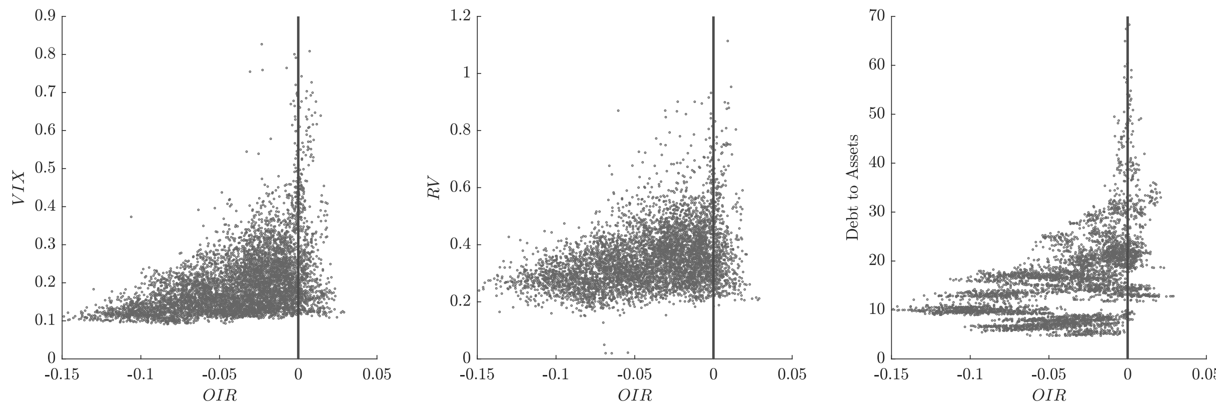
Brunnermeier and Pedersen (2009) and Adrian and Shin (2014) document that tight funding of the intermediary sector leads to market illiquidity which, in turn, results in higher volatility. I follow these insights and plot the relationship between option intermediary risk, volatility, and leverage in Figure 7. The scatter plots between OIR and the volatility proxies document a striking pattern. OptInts reduce their loss potential (increase in OIR) linearly with expected and realized volatility.[15] In fact, OIR is often positive when volatility is extremely high, indicating that OptInts sometimes profit from crashes. This is because OIR is positive only when the weighted positions of OptInts are net long. Recall that a byproduct of managing an option trading book is that OptInts are either net short or net long options. When OptInts are net short, they gain the options' time value but make a loss when volatility is high. When OptInts are net long, they pay the option's time value but profit from high volatility. Astonishingly, the P&L is often net short when volatility is very low and net long when volatility is peaking. In fact, the realized P&L is more often positive than negative for volatility levels above 50%.

B. Risk Transfer to Public

MMs manage positional risk smart and efficiently, and their market expectations are often very accurate. The question is how OptInts induce a risk transfer back to other market participants

[15] Table G.2 shows that the results are highly significant, do not change with a variety of controls, and also hold for the variance risk premium and left-tail volatility.

Figure 7. Option Intermediary Risk, Volatility, and Intermediary Constraints



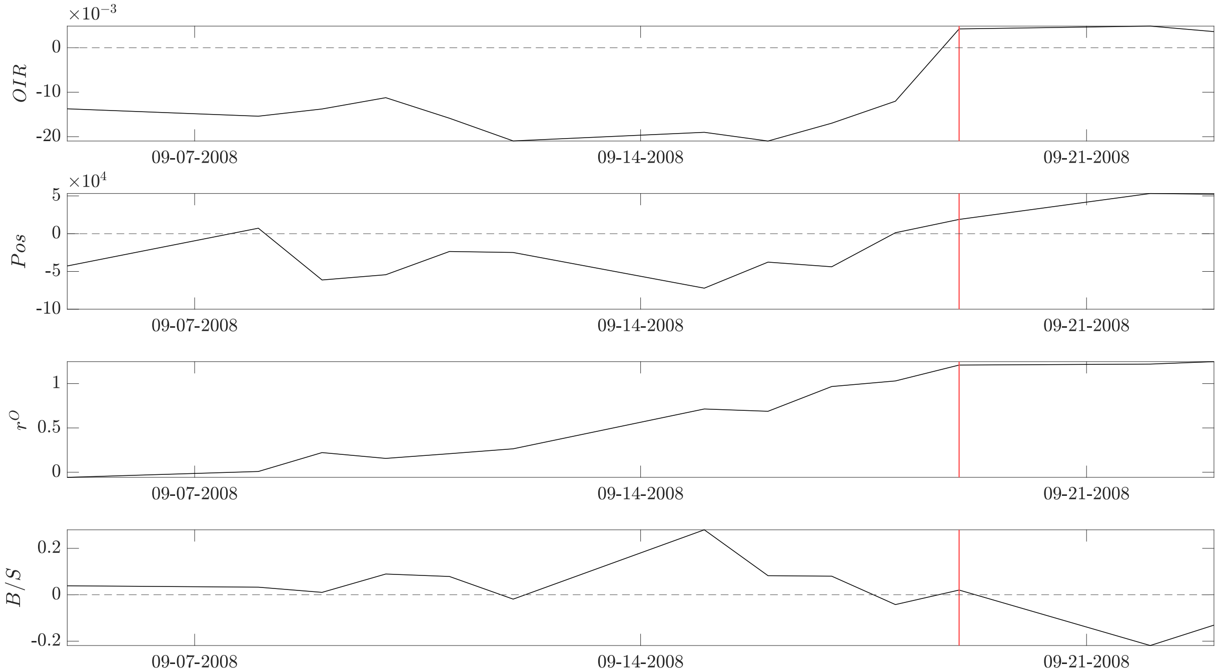
The figure plots the option intermediary risk OIR against the CBOE's volatility index VIX, realized volatility RV calculated from high-frequency returns, and the leverage ratio from He, Kelly, and Manela (2017).

whenever volatility is high. OptInts cannot simply close positions due to their role as market makers. Instead, they have to incentivize other investors to sell options so that intermediaries can reduce their short positions.

The mechanisms in order are visualized in Figure 8. The plot shows intermediary risk, net positions, the average option return, and the ratio between buy and sell spreads from Sep 5$^{\text{th}}$ until Sep 23$^{\text{rd}}$, 2008. The spreads are determined as the total spread income of OptInts divided by delta of the option. Hence, the spread is the paid spread per unit of exposure to the underlying. The period is characterized by extreme uncertainty because financial institutions suffered severe damage and Lehman Brothers filed for bankruptcy. Intermediaries do not want to hold short positions, and options prices increase significantly. The cumulated returns are 120% over two weeks. Buy spreads were up to 27% higher than sell spreads. The spike in the spread ratio coincides with buy activity from investors (positions of OptInts become more net short). Intermediaries want to prevent to hold more short positions. Higher prices and buy spreads incentivize other market participants to sell options, respectively, close existing long positions to lock in the profit. This is reflected in the positions of intermediaries as they turn from net short to net long on Sep 17$^{\text{th}}$. The intermediary risk factor follows two days later and turns from potential loss to profit in case of high volatility.

I provide two pieces of evidence showing that OptInts actively incentivizes other market participants to sell options. First, I regress option returns calculated from ask (buy) and bid (sell) prices on OIR for three different samples. The first is the full sample. The second is a low intermediation risk sample. As Figure 7 shows, times with low intermediary risk are when intermediaries likely want to reduce risky short positions. I condition the sample on $-0.05 <$

Figure 8. Event Study: September 2008



The figure shows the evolution of option intermediary risk (OIR), net intermediary positions (Pos), cumulated average option returns (r^O), and the ratio between effective buy and sell spreads per unit of exposure to the underlying (B/S) for the period Sep 5$^{\text{th}}$ to Sep 23$^{\text{rd}}$, 2008. The red line marks the day on which the scenario P&L of intermediaries in case of a -10% crash turned from loss to profit.

$m_t^{\text{P\&L}}$. The third is when OIR is near zero and volatility is above average because it is even more likely that intermediaries do not want to hold risky short positions when volatility is high. I use VIX levels above 20%. The results are given in Table VI.

Panel A reports results for the lag identification. The unconditional coefficient for bid (ask) option returns is -1.701% (-1.685%). Both coefficients are very similar, and the difference is less than 1%. Once I repeat the regression for states where OptInts most likely want to balance their positions towards zero, both coefficients diverge more. The bid coefficient equals 1.353% while the ask coefficient equals -1.290%, reflecting a difference of almost 5%. Next, I condition not only on OIR but also on VIX levels above its sample average of 20%. Besides the sharp increase in economic significance for the coefficients of OIR, the difference between the bid and ask coefficient is now larger than 13%. Panel B shows that the differences are even more pronounced using instrumental variable identification. While the unconditional coefficients differ by 6%, the conditional coefficients diverge significantly. The difference in the high volatility sample equals more than 65%. This supports the conclusion that OptInts increase bid quotes more to induce public sell orders. I report the coefficients for the controls in Table G.5. The results show that

Table VI. Ask and Bid Option Returns

	Full Sample		Low Intermediation Risk		VIX > 20%	
	r_t^{bid}	r_t^{ask}	r_t^{bid}	r_t^{ask}	r_t^{bid}	r_t^{ask}
Panel A: Lag Identification						
OIR	-1.701^{***}	-1.685^{***}	-1.353^{***}	-1.290^{***}	-4.677^{***}	-4.129^{***}
	(0.496)	(0.432)	(0.404)	(0.319)	(1.197)	(1.204)
adj. R^2 w/o	2.91	3.73	3.17	4.09	4.90	6.35
Panel B: Instrumental Variables						
OIR	-0.363^{***}	-0.342^{***}	-0.332^{**}	-0.242^{***}	-0.388^{***}	-0.232^{***}
	(0.116)	(0.097)	(0.137)	(0.060)	(0.124)	(0.074)
adj. R^2 w/o	1.78	2.74	2.28	3.54	4.21	5.95
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Entity FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
N	132,460	132,460	67,727	67,736	27,858	27,820

The table reports results of panel regression of end-of-day bid and ask option returns on option intermediary risk OIR and several control variables defined in Table IV. Standard errors (in parentheses) are clustered by entity. ***, **, and * denote significance at 1%, 5%, and 10%. The sample period is from Jan 2004 until Sep 2020. All independent variables are standardized and multiplied by 100. The top and bottom 0.5% outliers the returns are removed. The regression constant is not reported.

the volatility index VIX, the variance risk premium, and market returns also asymmetrically influence bid and ask quotes. Turbulent periods are characterized by intermediaries increasing sell prices more than buy prices such that other participants are encouraged to sell options so that intermediaries can reduce short positions. The higher explained variations also support the claim that the path of option premia is more predictable in such periods.

I find a similar pattern for effective spreads. I use the effective IV spread from Equation (18) but distinguish between buyer- or seller-initiated trades. I analyze the spreads for the same subsamples. Results are presented in Table G.6. The difference between the coefficients for OIR is significantly larger for the second subsample. The relationship between buy-initiated trades and OIR is -0.44 while the relationship is -0.33 for sell-initiated trades. Hence, when OIR is negative, volatility is high, and OptInts most likely want to balance positions, they charge one-third higher spreads when OIR decreases by one standard deviation. The difference between buyer- and seller-initiated spread income per unit of exposure to the underlying points in the same direction. The difference is 2.0% when OptInts most likely want to balance positions and 0.8% otherwise. Hence, buying becomes more expensive than it usually is. Interestingly, the widening gap is driven by call options. The results for spreads imply that OptInts require significantly higher premia for providing liquidity for buy transactions. The asymmetric bid quoting suggests that intermediaries actively want to induce public sell orders. OptInts are willing to pay higher prices so that public investors become insurers (short). The intensification transfers risk from intermediaries to other market participants because it helps OptInts to balance their positions. The risk transfer has been successful in the past (Figure 7), and intermediaries' option trading units have gained from past crashes.

VI. OPTION PRICING MODEL

In this section, I study if option intermediary risk drives crash risk premia. I design a continuous-time no-arbitrage model that characterizes the dynamics of the option surface, stock market volatility, and the OIR through a low-dimensional state vector. In contrast to most existing option pricing models, which put little restriction on the dynamics of the state vector, I use option intermediary risk as an observable state variable for the model estimation. This approach is very restrictive, however, sheds light on the importance of OIR for risk premia dynamics.

A. Motivation

The previous findings document that intermediaries reduce positions rapidly when volatility is high but are not averse otherwise. This speaks in favor of a time-varying appetite to intermediate the market and that the variation is related to high volatility states. Bates (2008) emphasizes the importance of heterogeneous agents with different attitudes toward crash risk for option markets. Chen, Joslin, and Ni (2019) extend the framework and consider a standard economy with intermediaries. The intermediaries have CRRA utility with a time-varying part:

$$U = E_0 = \left[\int_0^\infty e^{-\delta t} \frac{C^{1-\gamma}}{1-\gamma} e^{-\sum_0^{N_t} (\alpha_{\tau(n)} - \bar{\alpha})} dt \right], \tag{20}$$

where δ are time preferences, C the consumption share, γ the risk aversion, and α_t a time-varying variable that represents the ability to intermediate crash risk. The time variation can be interpreted as a VaR constraint that depends on the current state of the economy. Adrian and Shin (2010, 2014) show that intermediaries manage their leverage to keep the VaR to equity ratio constant. The effective risk aversion of intermediaries is

$$\gamma_t = \gamma - \frac{\alpha_{\tau(n)} - \bar{\alpha}}{\ln(C_{\tau(n)})}. \tag{21}$$

If the intermediating capacity is a function of crash risk $\alpha_{\tau(n)} = -ac_t$, the effective risk aversion of the intermediary will rise when the probability of jumps increases. This implies that with increased crash risk, intermediaries supply fewer options and instead demand options. This is only possible if the premium for crash risk increases. The effect should be most substantial for out-of-the-money put options because they have the highest loading on crash risk.

B. Model Dynamics

I use this motivation for a reduced-form option pricing model. If intermediaries have a time-varying aversion towards crash risk, *OIR* will be informative about priced crash risk implied option prices. Therefore, I use a model in which the jump intensity is conditioned on option intermediary risk. I fix a probability space $(\Omega,\ \mathcal{F},\ \mathbb{Q})$ and an information filtration $(\mathcal{F}_t)$. The filtration is assumed to satisfy the usual conditions. I specify the model under the equivalent martingale measure $\mathbb{Q}$ associated with risk-neutral probabilities. I let S_t denote the price of the stock index at time t and assume the following data-generating process:

$$\begin{aligned}
\frac{dS_t}{S_{t-}} &= (r_f - d)dt + \sqrt{V_{1,t}}dW^{\mathbb{Q}}_{1,t} + \sqrt{V_{2,t}}dW^{\mathbb{Q}}_{2,t} + \int_{\mathbb{R}\times\mathbb{R}^+} (e^x - 1)\tilde{\mu}^{\mathbb{Q}}(dt, dx), \\
dV_{1,t} &= \kappa_1(\overline{V_1} - V_{1,t})dt + \sigma_1\sqrt{V_{1,t}}dB^{\mathbb{Q}}_{1,t}, \\
dV_{2,t} &= \kappa_2(\overline{V_2} - V_{2,t})dt + \sigma_2\sqrt{V_{2,t}}dB^{\mathbb{Q}}_{2,t}, \\
dm_t &= -\kappa_m m_t dt + \sigma_m\sqrt{m_t}dB^{\mathbb{Q}}_{m,t} + \varrho\int_{\mathbb{R}\times\mathbb{R}^+} 1_{\{x<0\}}\, x\, \mu(dt, dx),
\end{aligned}$$

where $V_{1,t}$ is the first volatility factor, $V_{2,t}$ the second volatility factor, and m the option intermediary risk factor. For the five-dimensional Brownian motion $(dW^{\mathbb{Q}}_{1,t}, dW^{\mathbb{Q}}_{2,t}dB^{\mathbb{Q}}_{1,t}, dB^{\mathbb{Q}}_{2,t}, dB^{\mathbb{Q}}_{m,t})$, I impose that shocks between the index price and volatility factors are correlated with $\mathrm{corr}(dW^{\mathbb{Q}}_{1,t}, dB^{\mathbb{Q}}_{1,t})$ $= \rho_1$ and $\mathrm{corr}(dW^{\mathbb{Q}}_{2,t}, dB^{\mathbb{Q}}_{2,t}) = \rho_2$. The drift of S under the risk-neutral measure is the difference between the risk-free rate r_f and dividends d.

The integer-valued jump measure $\mu(dt, dx)$ counts the number of jumps. The jump compensator is given by $v^{\mathbb{Q}}_t$, and the difference $\tilde{\mu}^{\mathbb{Q}} = \mu(dt, dx) - v^{\mathbb{Q}}_t dt$ constitutes the martingale jump measure. That is, it adjusts the drift of the stock price by the expected jump contribution to satisfy the martingale property. For the jump-compensator, I assume a double exponential distribution for jumps in S_t. Andersen, Fusari, and Todorov (2015b) and Bardgett, Gourier, and Leippold (2019) use a similar specification and show that exponential jump distributions are superior to Gaussian jumps. The jump compensator reads as

$$\frac{\tilde{\mu}^{\mathbb{Q}}(dx)}{dx} = c^d_t \cdot 1_{\{x<0\}}\lambda^d\, e^{-\lambda^d|x|} + c^u_t \cdot 1_{\{x>0\}}\lambda^u\, e^{-\lambda^u x}.$$

$\lambda^{d,u}$ are the tail rate parameters so that the expected jump sizes equal $1/\lambda^{d,u}$. The tails of the distribution decay faster for higher values of the parameters. That is, higher values are associated with less thick tails. The jumps in m_t are conditioned on negative jumps in the underlying. The jump size is scaled by the parameter ϱ. The jump sizes are time-invariant so that the model is within the affine framework of Duffie, Pan, and Singleton (2000). The time variations are induced by the jump arrival intensities $c_t^{d,u}$. My specification allows differentiating between the arrival of negative and positive jumps. I specify the jump intensities as

$$c_t^{d,u} = c_m^{d,u} m_t + c_1^{d,u} V_{1,t} + c_2^{d,u} V_{2,t}.$$

The model falls under the class of self-exciting jump models because jumps in m make future jumps in m and S more likely, allowing for direct feedback between realized jumps, future jumps, and the positional risk of option intermediaries.

I model the dynamics of intermediary risk as a CIR process which implies that the variable is strictly positive. This requires a transformation of my empirical factor. I use this approach because times of increased jump activity are associated with OIR being close to zero, implying that intermediaries have little risk exposure in such states. It is hard for a model to pick up this particular feature. Therefore, I simplify my estimation process by projecting OIR on the real positive line $\mathbb{R}^+$. I add one to the empirical factor and further standardize the variable. The correlation between the empirical factor and the transformed factor equals one. The model's characteristic function is discussed in Appendix C and the pricing method in Appendix D.

C. Model Discussion

Option pricing models characterize the dynamics of the equity-index option surface through factors determining the volatility of the underlying stock market. The proposed model assumes a three-factor structure. Single volatility factor models provide a good fit to the option surface (Eraker, 2004; Broadie, Chernov, and Johannes, 2007) but exhibit some fundamental problems. The models do not allow for independent fluctuations between volatility levels, the slope (volatility smile), and the term structure of volatility. Christoffersen, Heston, and Jacobs (2009), among others, show that at least two factors are necessary to provide an adequate description of volatility. More recently, Andersen, Fusari, and Todorov (2015b) show that regular volatility factors cannot span the dynamics of priced downside risk and highlight that a third factor that accounts for fluctuations in the left tail is necessary.

The most notable feature of my model is the self-exciting jump process m, which is conditioned on the observable intermediary risk factor OIR. I assume that periods of increased jump activity are reflected by how option intermediaries manage their positions. A reduction of intermediaries' short positions (higher m) is associated with a higher crash risk premium. More specifically, a rise in m increases the probability of observing jumps, boosting all option prices. The choice implies that I do not account for the unconditional relationship between intermediary risk and option premia. Instead, I use the fact that intermediaries are concerned about risky positions in turbulent times, mirroring the finding that short positions and risks are transferred to the public. The dependency on OIR substantially restricts the model's degrees of freedom. Most option pricing models extract the latent state variables in a joint optimization over the parameter space and state realizations.[16] Because parameters are usually fixed, time variation in the option surface only depends on the state vector. I force m_t to exactly match the observable dynamics of OIR and the other two factors $V_{1,t}$ and $V_{2,t}$ to be in line with the dynamics of a nonparametric spot volatility estimator. None of the state variables can vary freely, and all have an empirical counterpart that is economically tractable. Therefore, the model's performance provides information about the imposed channel. A good performance indicates that intermediary risk is vital for option premia and that intermediaries' risk aversion varies in crash risk. I provide details on the estimation procedure in Appendix E.

D. Parameter Estimates

The annualized posterior means and standard deviations of the model parameters are reported in Table VII.

Table VII. Parameter Estimates

	Estimate	SD		Estimate	SD		Estimate	SD
$\overline{V_1}$	0.0334	0.0042	ρ_2	-0.8342	0.0099	c_1^d	23.8932	3.1792
κ_1	4.5284	0.1830	κ_m	26.7462	2.7543	c_2^d	35.2580	5.2703
σ_1	0.8572	0.0538	σ_m	0.7665	0.3634	c_m^u	0.2263	0.0113
ρ_1	-0.7097	0.0671	ϱ	17.7769	4.9029	c_1^u	16.2027	0.4934
$\overline{V_2}$	0.0089	0.0139	λ^d	16.9469	0.6109	c_2^u	14.6722	1.0722
κ_2	0.1048	0.0765	λ^u	41.2723	1.5017			
σ_2	0.8968	0.0403	c_m^d	0.4023	0.0197			

The table reports results for the estimated parameters using S&P 500 options. The data is sampled every Wednesday and spans the in-sample period from January 1998 until December 2017. All parameters are reported in annualized return units. The standard deviation is calculated from the accepted solutions of the MCMC sampler.

The first volatility factor corresponds to a short-term factor that mean-reverts relatively quickly. The long-term mean of the factor in terms of volatility equals 18.28%, and the variance of the factor is 0.86. Both estimates are relatively high because the model does not allow for jumps in volatility (Duffie, Pan, and Singleton, 2000). The second volatility factor corresponds to long-term volatility with a lower mean-reversion level. The mean-reversion speed is prolonged, indicating that the factor is persistent. The factor is also erratic. Noteworthy, the correlations

[16] Corsi, Fusari, and Vecchia (2013) and Christoffersen, Feunou, Jacobs, and Meddahi (2014) use realized variance constructed from high-frequency data as a direct input for option pricing models.

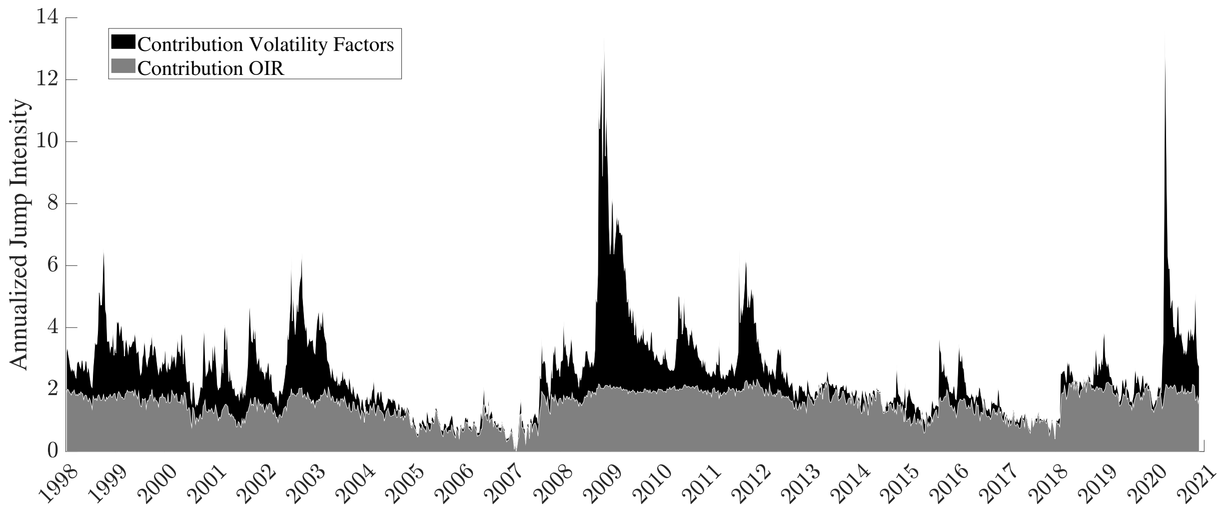
between the Brownian motions are significantly smaller than other models in which the estimation only relies on option data. The estimated jump size in case of a negative (positive) jump $1/\lambda^u$ equals 5.90% (2.42%) but are highly asymmetrically distributed. The impact of negative jumps is significantly more severe than that of positive jumps. The most interesting parameter corresponds to m. The information implied in option prices combined with the intermediary risk factor suggests a strong desire to balance positions towards zero, as implied by the high-mean reversion speed. The factor fluctuates a lot and has a high variance. If negative jumps occur, the jump size is scaled by $\varrho = 17.80$. This indicates that jumps are very likely to cluster, and the probability of subsequent jumps stays high after observing a jump. This translates to higher option premia over more extended periods. The loading on the jump intensity reveals interesting details. The jumps load is consistently stronger on intermediary risk than on volatility taking the level of the state variables into account. This is also true for positive jumps and seems reasonable from an intermediary risk perspective. Any volatility can be harmful to delta-hedged intermediaries.

E. Crash Risk Premia

The parameter estimates imply that the jump premium rises when intermediaries are reluctant to hold option positions. This boosts option prices and intends to induce public sell orders so that intermediaries can reduce short positions. The importance of the channel is visualized in Figure 9, which shows the negative jump arrival intensity split by contributions of *OIR* and volatility.

The plot highlights that priced crash risk depends on intermediary risk. The contribution equals almost two-thirds on average. Both volatility factors only contribute when markets are very turbulent such as during the financial or the Corona crisis. Interestingly, the contribution of the intermediary risk factor often rises before the contribution of volatility increases. This explains why jump premia move independently from volatility (Bollerslev and Todorov, 2011). The plot also highlights that the intermediary risk factor takes longer to recover after crises. The contribution stayed elevated after the financial crisis and the bust of a volatility fund in early 2018. This is in line with the evidence of Jackwerth (2000) and Andersen, Fusari, and Todorov (2020). Priced crash risk stays elevated much longer than volatility after a crisis. I exemplify how a shock to intermediary risk affects option prices in Figure 10. The plot shows short-term put prices and implied volatility after a one standard deviation shock in intermediary risk keeping everything else equal. The shock increases the prices of all options but has the highest absolute effect on deep out-of-the-money puts. IVs increase by approx. 10%, translating to a two percentage points increase for ATM IVs and seven percentage points for OTM IVs. In terms of prices, the shock increases the deepest OTM put by more than 100% and the ATM contract

Figure 9. Time Variation in Negative Crash Risk



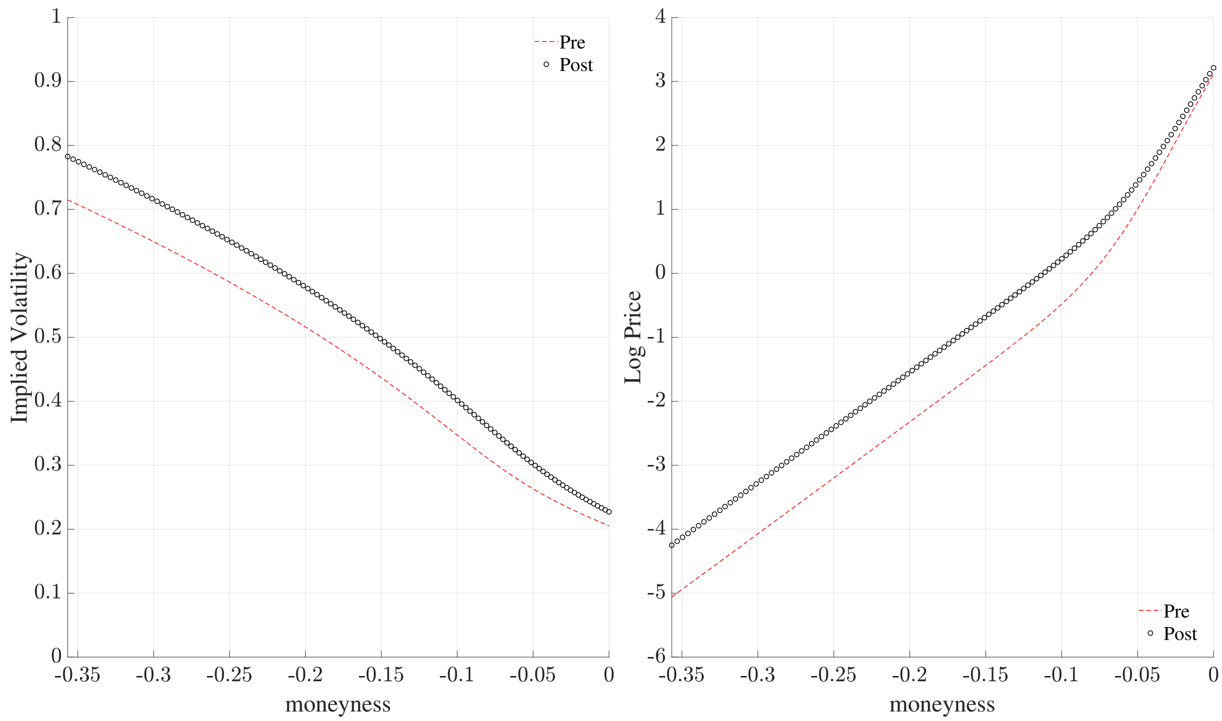
The figure shows the negative jump intensity $c_t^d = c_m^d m_t + c_1^d V_1 + c_2^d V_2$ using the estimated parameters and state variables. The jump intensity reflects the expected number of jumps per year. The grey area is the contribution from the intermediary risk factor, while the black area is the contribution from both volatility factors.

by 10%. A shock to intermediary constraints has more impact on out-of-the-money puts. The options are hard to hedge because of their little directional exposure. However, once the market jumps downwards, the value increases rapidly, and the option seller must make payments that exceed the initial premium. If intermediaries are more constrained because the probability of crashes increases, they will supply fewer OTM puts. By increasing the prices relative to other options, intermediaries incentivize other market participants to become option sellers or close existing short positions to lock in profits.

F. Pricing Performance

Figure G.4 plots the estimated trajectories of the latent state variable and the total RMSE. Panel A shows that the model-implied variance fits the high-frequency estimate well. Some deviations occur around the dot-com bubble and financial crisis. The model-implied values seem to have higher levels after turbulent times, but the dynamics are very similar. The fit in quiet times such as 2004-2008 or 2013-2017 is almost perfect. Panel B shows that the RMSE is low. The average is 0.5%, and the max equals 3%. The peaks in the RMSE coincide with the periods in which the volatility fit is also worse. Figure 11 visualizes the fit to the option surface over time. The fit to short-term characteristics, such as the 30-day ATM IV level and IV skewness, is quite satisfactory. The model underestimates skewness in turbulent times. However, the problem of higher errors during turbulent times is a feature that almost all option pricing models share.

Figure 10. Impact of OIR Shock on Put Options

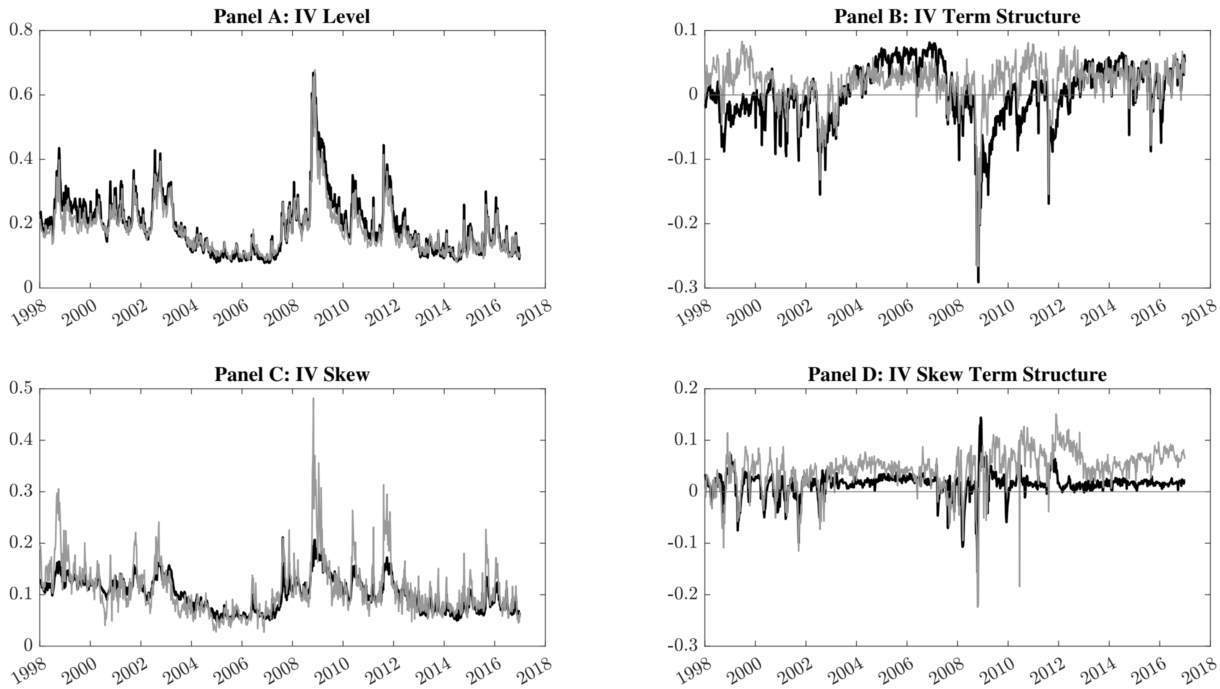


The figure shows changes in implied volatilities and log prices for short-term put options before and after a one-standard deviation shock in option intermediary risk OIR. The instantaneous volatility for both days is set at 20% and maturity is 7 days.

Turning to the term structure, I observe that the model has problems fitting the level of the term structure after turbulent times. The model-implied option prices underestimate long-term options in the aftermath of the 1998 Russian crisis and after the 2008 financial crisis. The most challenging feature to fit is the IV skew term structure. The fit is relatively well in turbulent times. The dynamics are well captured at the beginning of the sample and during the financial crisis. However, there are some significant outliers after the financial crisis. Surprisingly, the fit seems poorer at calmer periods where the skew term structure seems consistently higher than the model-implied.

The out-of-sample pricing performance is described in Appendix F and Table F.1. The results show that the pricing performance is generally outstanding. The RMSE is very low across the whole option surface. The fit to the deep-OTM options is quite remarkable, with an error below 2%, both for short-term and long-term options. The fit is also very good for short-term OTM calls. The intermediary factor seems to provide information for the pricing of both sides of the return distribution. The pricing performance is also superior for every region of the option

Figure 11. Fit to Option Surface Characteristics



The figure shows the model-implied (black line) and actual (grey line) option surface characteristics. Panel A shows the IV level of a short-term ATM contract. Panel B plots the IV skewness defined as the IV difference of a short-term OTM put and OTM call. Panel C displays the IV term structure defined as the ATM IV difference of a long- and short-term options. Panel D shows the IV skew term structure defined as the skewness difference of long-term and short-term options. OTM puts have moneyness $b = -2$, OTM calls have moneyness $b = 2$. Short-term options have 30 days maturity and long-term options 300 days to maturity.

surface. The intermediary factor provides a performance improvement of at least 10%, except for long-term ATM options. For instance, the short-term fit for the one-factor model is approx. 40% worse across all options. Nevertheless, both benchmark models perform relatively well.

VII. CONCLUSION

This paper shows that liquidity and risk premia in the index option market vary with option intermediaries' positions. I show that higher option intermediary risk – larger losses conditional on a market crash – causes higher risk premia and wider option spreads. Intermediaries require more compensation for bearing risks and simultaneously want to reduce subsequent option demand. I document that option intermediary risk is low when crash risk is high. Intermediaries asymmetrically adjust option quotes and spreads to incentivize sell orders. The risks associated with their option positions are transferred to other investors. I estimate an option pricing model and find that intermediaries' reluctance to hold risky option positions drives variation in crash risk premia independent of volatility.

My results suggest that many puzzles about option-implied risk premia relate to intermediaries' management of option positions. Models that target risk premia require heterogeneous agents and should allow for interactions between prices and positions. My results imply that intermediaries' risk tolerance is low around market turmoils, resulting in a reluctance to hold option positions. Less funding, fewer intermediation capacities, updated risk models, or tighter regulatory requirements likely affect intermediaries' ability to trade options. However, my analysis suggests that the option intermediary sector does not necessarily need stricter regulatory requirements. Intermediaries rapidly reduce positions before market turmoils, suggesting they are proficient at predicting crashes and adjusting their risks and positions accordingly. Higher capital requirements may artificially decrease the liquidity provision of intermediaries and affect the risk-sharing function of the option market.

REFERENCES

Adrian, T., E. Etula, and T. Muir, 2014, Financial intermediaries and the cross-section of asset returns, *Journal of Finance* 69, 2557–2596.

Adrian, T., and H. Shin, 2010, Liquidity and leverage, *Journal of Financial Intermediation* 19, 418–437.

Adrian, T., and H. Shin, 2014, Procyclical leverage and value-at-risk, *Review of Financial Studies* 27, 373–403.

Almeida, C., and G. Freire, 2022, Demand in the Option Market and the Pricing Kernel, *Working Paper*.

Amihud, Y., 2002, Illiquidity and stock returns: cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.

Amihud, Y., and H. Mendelson, 1980, Dealership market: Market-making with inventory, *Journal of Financial Economics* 8, 31–53.

Andersen, T., I. Archakov, L. Grund, N. Hautsch, Y. Li, S. Nasekin, I. Nolte, M. C. Pham, S. Taylor, and V. Todorov, 2021, A Descriptive Study of High-Frequency Trade and Quote Option Data*, *The Journal of Financial Econometrics* 19, 128–177.

Andersen, T., L. Benzoni, and J. Lund, 2002, An Empirical Investigation of Continuous-Time Equity Return Models, *Journal of Finance* 57, 1239–1284.

Andersen, T., O. Bondarenko, and M. T. Gonzalez-Perez, 2015, Exploring Return Dynamics via Corridor Implied Volatility, *Review of Financial Studies* 28, 2902–2945.

Andersen, T., N. Fusari, and V. Todorov, 2015a, Parametric Inference and Dynamic State Recovery From Option Panels, *Econometrica* 83, 1081–1145.

Andersen, T., N. Fusari, and V. Todorov, 2015b, The risk premia embedded in index options, *Journal of Financial Economics* 117, 558–584.

Andersen, T., N. Fusari, and V. Todorov, 2020, The Pricing of Tail Risk and the Equity Premium: Evidence From International Option Markets, *Journal of Business & Economic Statistics* 38, 662–678.

Baird, A., 1992, *Option Market Making: Trading and Risk Analysis for the Financial and Commodity Option Markets*. (Wiley Finance Editions New York, NY).

Bakshi, G., and N. Kapadia, 2003, Delta-Hedged Gains and the Negative Market Volatility Risk Premium, *Review of Financial Studies* 16, 527–566.

Bank for International Settlements, 2019, Minimum capital requirements for market risk, .

Bardgett, C., E. Gourier, and M. Leippold, 2019, Inferring volatility dynamics and risk premia from the S&P 500 and VIX markets, *Journal of Financial Economics* 131, 593–618.

Barras, L., and M. Aytek, 2016, Does variance risk have two prices? Evidence from the equity and option markets, *Journal of Financial Economics* 121, 79–92.

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

Bates, D. S., 2003, Empirical option pricing: A retrospection, *Journal of Econometrics* 116, 387–404.

Bates, D. S., 2008, The market for crash risk, *Journal of Economic Dynamics and Control* 32, 2291–2321.

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

Beason, T., and D. Schreindorfer, 2022, Dissecting the Equity Premium, *Journal of Political Economy* 130, 2203–2222.

Bekaert, G., and M. Hoerova, 2014, The VIX, the variance premium and stock market volatility, *Journal of Econometrics* 183, 181–192.

Bergomi, L., 2016, *Stochastic Volatility Modeling.* (CRC Press Boca Raton, FL).

Black, F., and M. Scholes, 1973, The Pricing of Options and Corporate Liabilities, *Journal of Political Economy* 81, 637–654.

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

Bollerslev, T., and V. Todorov, 2011, Tails, Fears, and Risk Premia, *Journal of Finance* 66, 2165–2211.

Bollerslev, T., V. Todorov, and L. Xu, 2015, Tail risk premia and return predictability, *Journal of Financial Economics* 118, 113–134.

Broadie, M., M. Chernov, and M. Johannes, 2007, Model Specification and Risk Premia: Evidence from Futures Options, *Journal of Finance* 62, 1453–1490.

Brunnermeier, M., and L. Pedersen, 2009, Market Liquidity and Funding Liquidity, *Review of Financial Studies* 22, 2201–2238.

Brunnermeier, M., and Y. Sannikov, 2014, A Macroeconomic Model with a Financial Sector, *American Economic Review* 104, 379–421.

Carr, P., and D. B. Madan, 1999, Option valuation using the fast Fourier transform, *Journal of Computational Finance* 2, 61–73.

Carr, P., and L. Wu, 2020, Option Profit and Loss Attribution and Pricing: A New Framework, *Journal of Finance* 75, 2271–2316.

Chen, H., S. Joslin, and S. X. Ni, 2019, Demand for crash insurance, intermediary constraints, and risk premia in financial markets, *Review of Financial Studies* 32, 228–265.

Cheng, I.-H., 2019, The VIX premium, *Review of Financial Studies* 32, 180–227.

Chordia, T., and A. Subrahmanyam, 2004, Order imbalance and individual stock returns: Theory and evidence, *Journal of Financial Economics* 72, 485–518.

Christoffersen, P., B. Feunou, K. Jacobs, and N. Meddahi, 2014, The Economic Value of Realized Volatility: Using High-Frequency Returns for Option Valuation, *Journal of Financial and Quantitative Analysis* 49, 663–697.

Christoffersen, P., R. Goyenko, K. Jacobs, and M. Karoui, 2018, Illiquidity premia in the equity options market, *Review of Financial Studies* 31, 811–851.

Christoffersen, P., S. Heston, and K. Jacobs, 2009, The Shape and Term Structure of the Index Option Smirk: Why Multifactor Stochastic Volatility Models Work So Well, *Management Science* 55, 1914–1932.

Christoffersen, P., and K. Jacobs, 2004, The importance of the loss function in option valuation, *Journal of Financial Economics* 72, 291–318.

Corsi, F., N. Fusari, and D. L. Vecchia, 2013, Realizing smiles: Options pricing with realized volatility, *Journal of Financial Economics* 107, 284–304.

Culp, C. L., Y. Nozawa, and P. Veronesi, 2018, Option-Based Credit Spreads, *American Economic Review* 108, 454–88.

Daglish, T., J. Hull, and W. Suo, 2007, Volatility surfaces: theory, rules of thumb, and empirical evidence, *Quantitative Finance* 7, 507–524.

Du, W., B. Hébert, and A. W. Huber, 2022, Are Intermediary Constraints Priced?, *Review of Financial Studies*.

Duffie, D., J. Pan, and K. Singleton, 2000, Transform Analysis and Asset Pricing for Affine Jump-Diffusions, *Econometrica* 68, 1343–1376.

Eraker, B., 2004, Do Stock Prices and Volatility Jump? Reconciling Evidence from Spot and Option Prices, *Journal of Finance* 59, 1367–1403.

Fang, F., and C. W. Oosterlee, 2009, A Novel Pricing Method for European Options Based on Fourier-Cosine Series Expansions, *SIAM Journal on Scientific Computing* 31, 826–848.

Fournier, M., and K. Jacobs, 2020, A tractable framework for option pricing with dynamic market maker inventory and wealth, *Journal of Financial and Quantitative Analysis* 55, 1117–1162.

Gabaix, X., 2012, Variable Rare Disasters: An Exactly Solved Framework for Ten Puzzles in Macro-Finance , *The Quarterly Journal of Economics* 127, 645–700.

Gârleanu, N., L. H. Pedersen, and A. M. Poteshman, 2009, Demand-based option pricing, *Review of Financial Studies* 22, 4259–4299.

Goyenko, R., and C. Zhang, 2020, Price Pressures and Noise in Option Returns, *Working Paper.*

Haddad, V., and T. Muir, 2021, Do intermediaries matter for aggregate asset prices?, *Journal of Finance* 76, 2719–2761.

He, Z., B. Kelly, and A. Manela, 2017, Intermediary asset pricing: New evidence from many asset classes, *Journal of Financial Economics* 126, 1–35.

He, Z., and A. Krishnamurthy, 2013, Intermediary Asset Pricing, *American Economic Review* 103, 732–70.

Hendershott, T., and M. Seasholes, 2007, Market Maker Inventories and Stock Prices, *American Economic Review* 97, 210–214.

Heston, S., 1993, A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options, *Review of Financial Studies* 6, 327–43.

Ho, T., and H. Stoll, 1981, Optimal dealer pricing under transactions and return uncertainty, *Journal of Financial Economics* 9, 47–73.

Ho, T. S., and R. G. Macris, 1984, Dealer bid-ask quotes and transaction prices: An empirical study of some AMEX options, *Journal of Finance* 39, 23–45.

Hogg, D. W., and D. Foreman-Mackey, 2018, Data Analysis Recipes: Using Markov Chain Monte Carlo, *The Astrophysical Journal Supplement Series* 236, 11.

Hu, G. X., J. Pan, and J. Wang, 2013, Noise as Information for Illiquidity, *Journal of Finance* 68, 2341–2382.

Huang, J., and L. Wu, 2004, Specification Analysis of Option Pricing Models Based on Time-Changed Lévy Processes, *Journal of Finance* 59, 1405–1439.

Hull, J., 2009, *Options, Futures, and Other Derivatives.* (Pearson Prentice Hall, New York).

Jackwerth, J. C., 2000, Recovering Risk Aversion from Option Prices and Realized Returns, *Review of Financial Studies* 13, 433–451.

Jacobs, K., A. T. Mai, and P. Pederzoli, 2021, Market-Maker Supply and Investor Demand for SPX Options: A VAR Approach, *Working Paper.*

Jiang, G., and Y. Tian, 2005, The Model-Free Implied Volatility and Its Information Content, *Review of Financial Studies* 18, 1305–1342.

Johannes, M. S., N. G. Polson, and J. R. Stroud, 2009, Optimal Filtering of Jump Diffusions: Extracting Latent States from Asset Prices, *Review of Financial Studies* 22, 2559–2599.

Kelly, B., and H. Jiang, 2014, Tail Risk and Asset Prices, *Review of Financial Studies* 27, 2841–2871.

Kondor, P., and D. Vayanos, 2019, Liquidity Risk and the Dynamics of Arbitrage Capital, *Journal of Finance* 74, 1139–1173.

Muir, T., 2017, Financial Crises and Risk Premia*, *The Quarterly Journal of Economics* 132, 765–809.

Muravyev, D., 2016, Order Flow and Expected Option Returns, *Journal of Finance* 71, 673–708.

Muravyev, D., and X. Ni, 2020, Why do option returns change sign from day to night?, *Journal of Financial Economics* 136, 219–238.

Newey, W., and K. West, 1987, A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* 55, 703–08.

Ni, S. X., N. D. Pearson, A. M. Poteshman, and J. White, 2021, Does Option Trading Have a Pervasive Impact on Underlying Stock Prices?, *Review of Financial Studies* 34, 1952–1986.

Pan, J., 2002, The jump-risk premia implicit in options: evidence from an integrated time-series study, *Journal of Financial Economics* 63, 3–50.

Sepp, A., 2012, An approximate distribution of delta-hedging errors in a jump-diffusion model with discrete trading and transaction costs, *Quantitative Finance* 12, 1119–1141.

Sepp, A., 2013, When You Hedge Discretely: Optimization of Sharpe Ratio for Delta-Hedging Strategy under Discrete Hedging and Transaction Costs, *Journal of Investment Strategies* 3, 19–59.

Zhou, H., 2018, Variance Risk Premia, Asset Predictability Puzzles, and Macroeconomic Uncertainty, *Annual Review of Financial Economics* 10, 481–497.

APPENDIX

A DATA

The volume data sometimes reports negative maturities and cuts of the first three digits of strikes exceeding 1000. To overcome the issues, I implement a simple yet effective correction for all options for which the matching with the quote data was unsuccessful. I create list with option chains that report a negative time-to-maturity and obtain the correct expiry date from option quotes. The correct expiry is often two weeks after the reported expiry and is found by comparing option characteristics such as strike, flag, and traded volume across the two data banks. The correction concerns 195 option chains and 86 unique expiry dates. Second, all unmatched options for which the strike price was more than 40% away from the current S&P 500 level, I add 1000 to the reported strike and compare characteristics of the adjusted option contract with options quote. The moneyness range for SPX options was relatively narrow during 1990 and early 2000, with options rarely trading below 75% moneyness. For July 1998, for instance, only 20% of all option contracts could be matched. The majority of the unmatched option contracts report strikes from 5 to 250, although the level of the S&P 500 was moving around 1150. It is extremely unlikely that 80% of all options trade in a moneyness region of 0.5% to 20%. Indeed, 99.32% of the data in July 1998 was matched after adjusting the strike prices. Figure G.1 shows that up 90% of the data would have been dismissed in certain months without applying a correction algorithm. Panel A highlights that errors in the expiry date occurred in the early 1990 and during the financial crisis. Adjusting the incorrect expiry date was especially important in the beginning of the samples because not many different time-to-maturities were available and the option chain would have been dismissed over the full life cycle of option. Panel B highlights that the mismatch induced by the incorrect strike prices is as high as 95% of all option contracts traded. The problem occurs during 1998 until 2005. After correcting both data reporting errors, the ratio of of volume data that was matched with option quotes was almost perfect except for certain months in 2000/2001 and 2003. Only 1.24% of the total traded volume could not be matched.

B SENSITIVITIES

The PDE of the Black and Scholes (1973) model allows to express all partial derivatives in terms of the second-order partial derivative with respect to the underlying Γ (Hull, 2009; Carr and Wu, 2020). The sensitivity of a call price with respect to time is given by

$$\Theta = -\frac{1}{2}\Gamma F_t^2 IV_t^2. \tag{22}$$

I express the remaining sensitivities not in terms of units but in terms of percentage changes ('cash' greeks). The cash vega is given by

$$\nu\, IV_t = IV_t^2 \tau \Gamma F_t^2, \tag{23}$$

where τ is the time-to-maturity. The second-order derivative with respect to IV is given by

$$\vartheta\, IV_t^2 = \Gamma F_t^2 z^- z^+, \tag{24}$$

where $z^{\pm} = \ln\left(\frac{K}{S_t}\right) \pm \frac{1}{2} IV_t^2 \tau$ is a convexity-adjusted measure of moneyness. Finally, the 'cross-sensitivity' w.r.t. changes in F_t and IV_t can be expressed as

$$\varepsilon\, IV_t\, F_t = \Gamma F^2 z^+. \tag{25}$$

C CHARACTERISTIC FUNCTION

Options give the right to buy (sell) the underlying in exchange for the strike price of an option. Options only pay off when the terminal value of the underlying is above (beyond) the strike price. Therefore, prices are expected, discounted payoffs. The conditional probability density function of the underlying's log-return y must be known to price an option. A call option is formally given by

$$C_{K,\tau} = e^{r_f \tau}\mathbb{E}^{\mathbb{Q}}\left[\max(S_T - K, 0)\right] = e^{r_f \tau} S_t \int_{\mathbb{R}} max(e^{y_T} - e^K, 0) f_{y_T}(y)dy. \tag{26}$$

The option price depends on the risk-free rate r_f, the strike price K, time-to-maturity $\tau = T - t$, the current value of the underlying S_t, and the conditional probability density function f_{y_T} of the return between t and T conditional on all available information at t.

Most option pricing models do not have a known moment-generating function that determines the form of the probability density function. However, as introduced by Heston (1993) and formally developed by Duffie, Pan, and Singleton (2000), the probability density function of affine jump diffusion processes can be determined by its generalized characteristic function.

$$\phi(\tau, y_t, V_t, \omega_t, \Gamma_t, u) = \mathbb{E}_t[e^{y_\tau u}] = \int_{-\infty}^{\infty} e^{uy} f_Y(y)dy, \tag{27}$$

where $u \in \mathbb{C}$ is the characteristic exponent. The model I analyze is given by

$$\begin{aligned}
\frac{dS_t}{S_{t-}} &= (r_f - d)dt + \sqrt{V_{1,t}}dW_{1,t}^{\mathbb{Q}} + \sqrt{V_{2,t}}dW_{2,t}^{\mathbb{Q}} + \int_{\mathbb{R}\times\mathbb{R}^+} (e^x - 1)\tilde{\mu}^{\mathbb{Q}}(dt, dx), \\
dV_{1,t} &= \kappa_1(\overline{V_1} - V_{1,t})dt + \sigma_1\sqrt{V_{1,t}}dB_{1,t}^{\mathbb{Q}}, \\
dV_{2,t} &= \kappa_2(\overline{V_2} - V_{2,t})dt + \sigma_2\sqrt{V_{2,t}}dB_{2,t}^{\mathbb{Q}}, \\
dm_t &= -\kappa_m m_t dt + \sigma_m\sqrt{m_t}dB_{m,t}^{\mathbb{Q}} + \varrho\int_{\mathbb{R}\times\mathbb{R}^+} 1_{\{x<0\}}\, x\, \mu(dt, dx).
\end{aligned}$$

For easier notation, I define dN_t as the Poisson process counting the jumps and Z as the jump sizes. The model is affine (Duffie, Pan, and Singleton, 2000) and the form of the generalized characteristic function can be expressed as

$$\phi(\tau, y_t, V_t, \omega_t, \Gamma_t, u) = \mathbb{E}_t[e^{y_\tau u}] := e^{\alpha(\tau,u)+\beta_1(\tau,u)V_{1,t}+\beta_2(\tau,u)V_{2,t}+\beta_m(\tau,u)m_t+uy_t}, \tag{28}$$

where $\tau = T - t$ is used for simplicity. The characteristic function is a stochastic process and a martingale. Therefore, we can apply Ito's Lemma and solve

$$\begin{aligned}&\frac{\partial\phi}{\partial\tau}+\frac{\partial\phi}{\partial y}\mu_{y^c,t}+\frac{\partial\phi}{\partial V_1}\mu_{1,t}+\frac{\partial\phi}{\partial V_2}\mu_{2,t}+\frac{\partial\phi}{\partial m}\mu_{m,t}+\frac{1}{2}\frac{\partial^2\phi}{\partial y^2}\sigma^2_{y^c,t}+\frac{1}{2}\frac{\partial^2\phi}{\partial V_1^2}\sigma^2_{V_1,t}+\frac{1}{2}\frac{\partial^2\phi}{\partial V_2^2}\sigma^2_{V_2,t}\\&+\frac{1}{2}\frac{\partial^2\phi}{\partial m^2}\sigma^2_{m,t}+\frac{\partial^2\phi}{\partial y\partial V_1}\sigma_{y^c,t}\sigma_{V_1,t}+\frac{\partial^2\phi}{\partial y\partial V_2}\sigma_{y^c,t}\sigma_{V_2,t}+\mathbb{E}[e^{uZ}-1]dN_t=0.\end{aligned} \tag{29}$$

Note that μ and σ denote the drift and variance of the respective process, y^c denotes the continuous part of y, and $\mathbb{E}[e^{uZ}-1]dN_t$ is the discontinuous part ot the characteristic function. The respective partial derivatives are given by

$$\begin{aligned}&\frac{\partial\phi}{\partial\tau}=\phi\left(-\frac{\partial\alpha}{\partial\tau}-\frac{\partial\beta_1}{\partial\tau}V_{1,t}-\frac{\partial\beta_2}{\partial\tau}V_{2,t}-\frac{\partial\beta_m}{\partial\tau}m\right),\\&\frac{\partial\phi}{\partial y}=\phi u,\qquad \frac{\partial^2\phi}{\partial y^2}=\phi u^2,\\&\frac{\partial\phi}{\partial V_1}=\phi\beta_1,\qquad \frac{\partial^2\phi}{\partial V_1^2}=\phi\beta_1^2,\end{aligned} \tag{30}$$

$$\begin{aligned}&\frac{\partial\phi}{\partial V_2}=\phi\beta_2,\qquad \frac{\partial^2\phi}{\partial V_2^2}=\phi\beta_2^2,\\&\frac{\partial\phi}{\partial m}=\phi\beta_m,\qquad \frac{\partial^2\phi}{\partial m^2}=\phi\beta_m^2,\\&\frac{\partial^2\phi}{\partial y\partial V_1}=\phi\beta_1 u,\qquad \frac{\partial^2\phi}{\partial y\partial V_2}=\phi\beta_2 u\end{aligned} \tag{31}$$

Plugging the partial derivatives into Equation (29), dividing by ϕ, and substituting dX_t with the drift and dX_t^2 with the variance of the respective process gives

$$\begin{aligned}&-\frac{\partial\alpha}{\partial\tau}-\frac{\partial\beta_1}{\partial\tau}V_{1,t}-\frac{\partial\beta_2}{\partial\tau}V_{2,t}-\frac{\partial\beta_m}{\partial\tau}m_t+u(r-\frac{1}{2}V_{1,t}-\frac{1}{2}V_{2,t}-\mathbb{E}\left[e^Z-1\right]\lambda_t)\\&+\beta_1\kappa_1(\theta_1-V_{1,t})+\beta_2\kappa_2(\theta_2-V_{2,t})+\beta_m[\kappa_m(\theta_m-m_t)]+\frac{1}{2}u^2(V_{1,t}+V_{2,t})\\&+\frac{1}{2}\beta_1^2\sigma_1^2V_{1,t}+\frac{1}{2}\beta_2^2\sigma_2^2V_{2,t}+\frac{1}{2}\beta_m^2\sigma_m^2m_t+\beta_1u\rho_1\sigma_1V_{1,t}+\beta_2u\rho_2\sigma_2V_{2,t}\\&+\mathbb{E}[e^{uZ}-1]\lambda_t=0.\end{aligned} \tag{32}$$

Equation (32) is equal to zero when the all terms multiplying the state variables equal zero and the constant term equals 0. Plugging in the jump intensity as well as sizes, and collecting the terms gives the system of ODEs. The boundary conditions are $\alpha(0)=\beta_1(0)=\beta_2(0)=\beta_m(0)=0$ so that

$$
\begin{aligned}
\frac{\partial \alpha}{\partial \tau} =& u(r-d) + \beta_1\kappa_1\theta_1 + \beta_2\kappa_2\theta_2 + \beta_m\kappa_m\theta_m \\
\frac{\partial \beta_1}{\partial \tau} =& -\frac{1}{2}(u-u^2) - \beta_1\kappa_1 + \frac{1}{2}\beta_1^2\sigma_1^2 + \beta_1 u\rho_1\sigma_1 - u\left(c_1^u J_u + c_1^d J_d\right) + c_1^u J_{u,m} + c_1^d J_{d,m}, \\
\frac{\partial \beta_2}{\partial \tau} =& -\frac{1}{2}(u-u^2) - \beta_2\kappa_2 + \frac{1}{2}\beta_2^2\sigma_2^2 + \beta_2 u\rho_2\sigma_2 - u\left(c_2^u J_u + c_2^d J_d\right) + c_2^u J_{u,m} + c_2^d J_{d,m}, \\
\frac{\partial \beta_m}{\partial \tau} =& -\beta_m\kappa_m + \frac{1}{2}\beta_m^2\sigma_m^2 - u\left(c_m^u J_u + c_m^d J_d\right) + c_m^u J_{u,m} + c_m^d J_{d,m}.
\end{aligned}
$$

where

$$
\begin{aligned}
J_u =& \frac{\lambda^u}{\lambda^u - 1} - 1, \quad J_d = \frac{\lambda^d}{\lambda^d + 1} - 1, \\
J_{u,m} =& \frac{\lambda^u}{u - \lambda^u} - 1, \quad J_{d,m} = \frac{\lambda^d}{u + \lambda^d - \beta_m \varrho} - 1.
\end{aligned}
$$

The self-exciting jump feature circumvents a closed-form solution, but the ODEs can be solved numerically.

D PRICING METHOD

I use the Fourier Cosine transform of Fang and Oosterlee (2009) to calculate European option prices. The method is computational faster and more accurate than other methods such as the Quadrature approach of Heston (1993) or the Fast Fourier Transform of Carr and Madan (1999). The price of a put option for state vector X_t, strike price K, and $y = \ln(\frac{S_T}{K})$ is given by

$$P_{K,\tau} \approx e^{r_f \tau} \sum_{j=0}^{M} a_j(X_t) \frac{2}{b-a} K(-\chi_j(a,0) + \psi_j(a,0)), \tag{33}$$

with

$$a_j(X_t) = \frac{2}{b-a} \mathbb{R} \left[\phi \left(\frac{j\pi}{b-a}, X_t \right) e^{ij\pi \frac{a}{b-a}} \right], \tag{34}$$

$$\chi_j(c,d) = \int_c^d e^y cos(j\pi \frac{y-a}{b-a}) dy, \tag{35}$$

$$\psi_j(c,d) = \int_c^d cos(j\pi \frac{y-a}{b-a}) dy. \tag{36}$$

The method approximates the integrals over the put's payoff and density functions by using the Euler formula. Because cosine series coefficients χ_j and ψ_j have analytical solutions, the method is computationally inexpensive.[17] The errors induced by truncating the integration domain from $[-\infty, \infty]$ to $[a, b]$ and by using a finite number of expansion terms M can be made very small. I follow Fang and Oosterlee (2009) in calculating a and b and use $M = 400$. I only determine put prices and use put-call parity to obtain the respective call prices.

[17] See Result 3.1 in Fang and Oosterlee (2009).

E ESTIMATION PROCEDURE

Implementing option pricing models requires jointly estimating the structural parameter vector θ and the latent state vector X_t. A common approach is to extract the state variables from index returns and option data using various filtering techniques. For example, Pan (2002) uses Generalized Method of Moments, Andersen, Benzoni, and Lund (2002) use Efficient Method of Moments, and Johannes, Polson, and Stroud (2009) use particle filters. I use an iterative approach proposed by Huang and Wu (2004) or Christoffersen, Heston, and Jacobs (2009) and combine it with Markov chain Monte Carlo (MCMC) sampling methods. More specifically, I first extract the latent state vector $X_t = [V_{1,t}, V_{2,t}]$ given an initial set of parameters θ and a loss function $L(X_t, \theta)$. Second, I treat the extracted state vector as given and extract the optimal parameters. I iterate between the two steps until no further improvement in the optimization problem is archived. As noted in Christoffersen, Heston, and Jacobs (2009), the procedure requires relatively few iterations to converge but potentially suffers from inconsistencies between the estimated state vector and parameters. I account for inconsistency by using MCMC sampling as a final optimization step. I use the quasi-optimal solution and sample over the parameters using a chain length of $10,000$. Hogg and Foreman-Mackey (2018) propose to use an optimizer in advance of sampling to ensure that the walker is not initialized at a completely irrelevant point. Conveniently, the sampling allows for obtaining a distribution of optimal parameters.

The choice of the loss function is well known to be critical for model estimation. Many studies use likelihood functions for parameter estimation and evaluate the model with its pricing error. Christoffersen and Jacobs (2004) show that inconsistencies in the loss function between model estimation and evaluation lead to suboptimal parameter estimates. Therefore, I follow Andersen, Fusari, and Todorov (2015a,b) and define the loss function in terms of pricing performance, that is

$$L\left(\{\hat{V_{1,t}}, \hat{V_{2,t}}\}_{t=1,\ldots,T}, \hat{\theta}\right) = \underset{\{X_t\}_{t=1,\ldots,T}, \theta \in \Omega}{\operatorname{argmin}} \quad \sum_{t=1}^{T} \left\{ \frac{\text{Option Fit}_t + \delta \times \text{Vol Fit}_t}{V_t^{ATM}} \right\}, \tag{37}$$

$$\text{Option Fit}_t = \frac{1}{N_t} \sum_{j=1}^{N_t} \left(IV_t(K, \tau) - IV(K, \tau, X_t, \theta) \right)^2,$$

$$\text{Vol Fit}_t = \left(\sqrt{\hat{V}_t} - \sqrt{V_(X_t, \theta)} \right)^2,$$

where V_t^{ATM} is the squared at-the-money Black-Scholes IV obtained from the shortest available maturity, IV_t is an option's observable BSM implied volatility, $IV(K, \tau, X_t, \delta)$ is the model-implied BSM implied volatility, $\hat{V}_t$ is a nonparametric diffusive spot variance estimator constructed from high-frequency returns, $V_(X_t, \theta) = V_{1,t} + V_{2,t}$ the model-implied diffusive spot variance, and δ a penalty term. Andersen, Fusari, and Todorov (2015a) formally develop the loss function and show that it ensures consistent estimation of the parameter and state vector. The estimator is a penalized nonlinear least square method consisting of two parts. Option Fit_t is the MSE in fitting observable and model-implied option prices. Vol Fit_t penalizes deviation of the model-implied variance from a nonparametric estimator, not only ensuring the no-arbitrage constraint between physical and risk-neutral variance but also regularizing the latent state variables. Without the penalization, $V_{1,t}$ and $V_{2,t}$ may in principle take any value. I use a penalty of $\delta = 0.2$.[18] Lastly, the standardization by V_t^{ATM} ensures that volatility states are weighted differently because the volatility of option pricing errors increases with the volatility.

I use the same data filters as in Andersen, Fusari, and Todorov (2015b) for the estimation. The filters are mild and remove illiquid in-the-money options, options with a time-to-maturity of fewer than seven days, options with zero bid price, and extreme out-of-the-money options. To obtain the parameter vector, I sample the data every Wednesday. This is standard in the literature (Bates, 2000; Christoffersen, Heston, and Jacobs, 2009; Andersen, Fusari, and Todorov, 2015b) and reduces the computational burden of my high-dimensional optimization problem. Note that I can still extract the state vector for every day once I obtain the optimal solution for the parameters.

Gârleanu, Pedersen, and Poteshman (2009) argue that the index option market most likely faced a structural break between 1996 and October 1997, going back to the introduction of S&P e-mini futures, futures options on the CME, Dow Jones options, and changes in margin requirements. Therefore, I set the start of my sample to January 7th, 1998. I divide the sample into an in-sample period (until 2017) and an out-of-sample period (from 2017 to 2021). The in-sample period contains 992 days, and the out-of-sample period 205. Finally, I use the Bipower variation from the Oxford-Man Realized Library as my nonparametric diffusive spot variance estimator. The data is available as of January 3rd, 2000. I fill the period prior to 2000 with the volatility estimates from the Risk Lab of Dacheng Xiu.[19]

[18] Andersen, Fusari, and Todorov (2015b) show that the estimation results do not depend on δ.

[19] See https://dachxiu.chicagobooth.edu/.

F OUT-OF-SAMPLE PERFORMANCE

I compare the performance of my model to two other models. The first model 1FGJ is the one-factor volatility model of Pan (2002) that features Gaussian jumps in S and a jump intensity $\lambda_t = \lambda_0 + \lambda_v V_t$. The second model 2FGJ is a two-factor model with Gaussian jumps and jump arrivals that evolve with $\lambda_t = \lambda_0 + \lambda_1 V_{1,t} + \lambda_2 V_{2,t}$. Both models are more parsimonious than my model but are not restricted to fit an observable variable. A three-factor model will most likely always outperform my model because the third variable can, in principle, take any value. The out-of-sample period is from 2017 until 2021 and contains more than 785,187 option quotes. The period also spans two major events for the option markets. The first is the "Volmageddon", the bust of a volatility fund in early 2018. The second is the Corona pandemic in 2020. The results are reported in Table F.1, which shows the RMSE $= \sqrt{\frac{1}{N_t}\sum_{j=1}^{N_t}\left(IV_t(K,\tau) - IV(K,\tau,X_t,\theta)\right)^2}$ for the self-exciting intermediary jump model and the relative performance of the other two models.

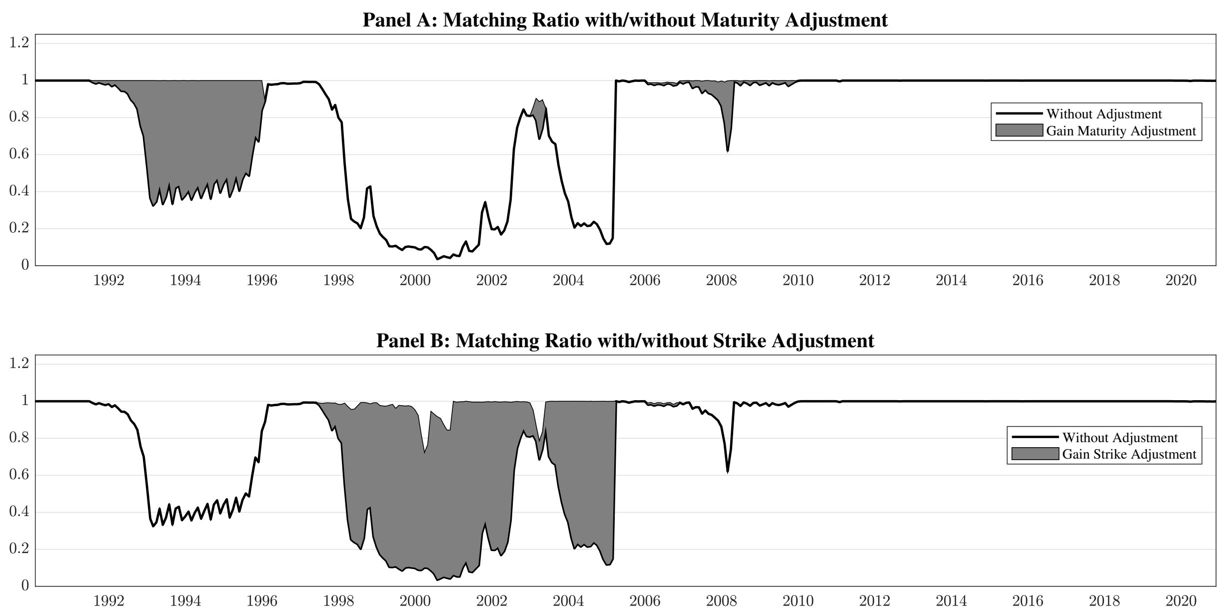
Table F.1. Out-of-Sample Pricing Performance

	RMSE in %		RRMSE 1FGJ		RRMSE 2FGJ	
	$\tau \leq 60$	$\tau > 60$	$\tau \leq 60$	$\tau > 60$	$\tau \leq 60$	$\tau > 60$
money ≤ -3	1.97	1.22	0.78	0.66	0.14	0.21
$-3 <$ money ≤ -1	0.90	0.64	0.54	0.41	0.12	0.34
$-1 <$ money ≤ 1	0.91	1.61	0.38	0.05	0.16	0.03
money > 1	0.65	1.78	0.41	0.38	0.28	0.15

The table reports median out-of-sample option RMSE for the self-exciting jump model and the ratio of a given model minus one (RRMSE). τ is the maturity and money is volatility- and maturity-adjusted moneyness defined as $\log(K/S)/IV_{ATM}\sqrt{\tau}$. The first bucket ≤ -3 are deep out-of-the-money puts, the second are out-of-the-money puts, the third are at-the-money options, and the fourth are out-of-money calls.

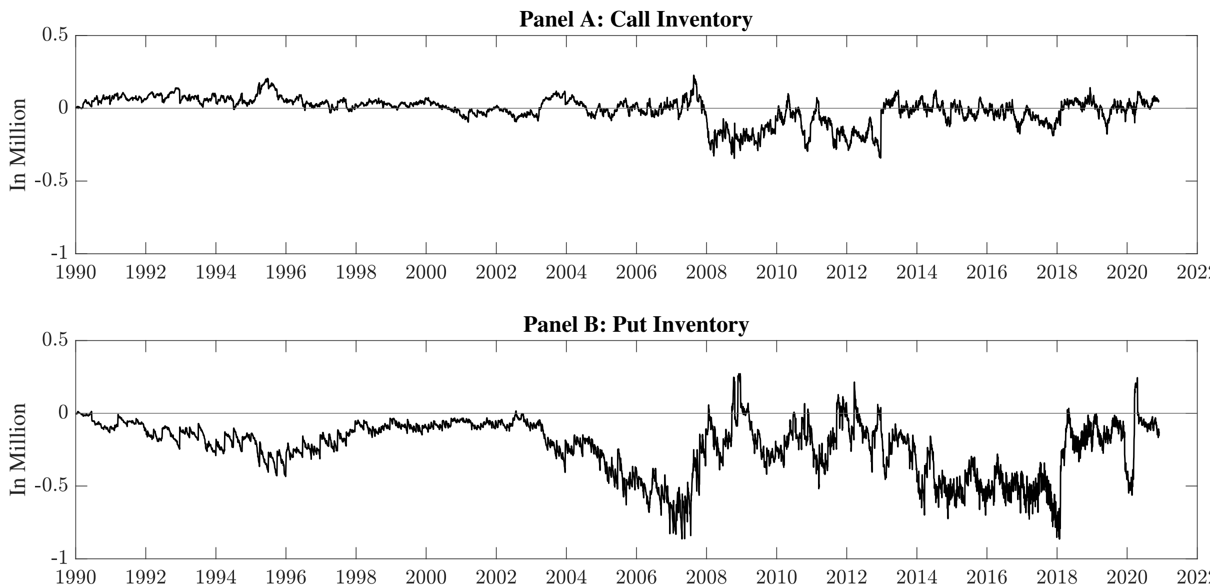
G ADDITIONAL FIGURES AND TABLES

Figure G.1. Recovery Ratio of Volume Data



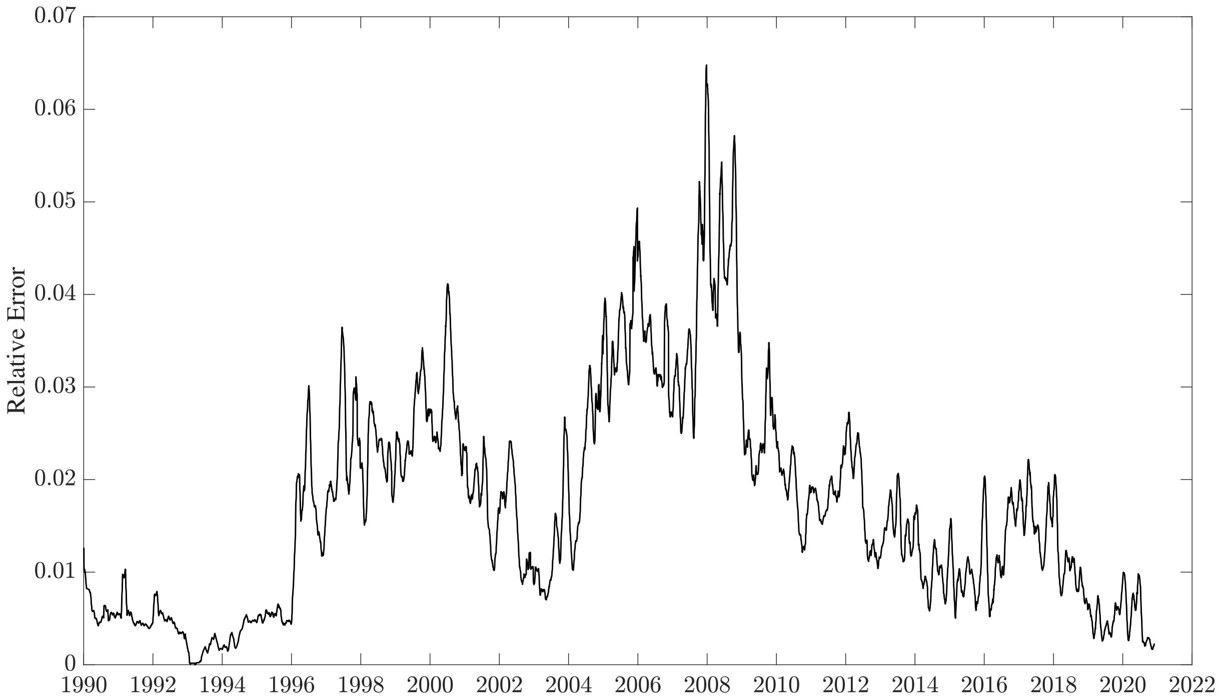
The figure shows the recovery ratio for the inventory data for 1990 until 2021. Panel A shows the ratio with/without maturity adjustment and Panel B depicts the ratio with/without strike adjustment.

Figure G.2. Time-Series of Put and Call Positions



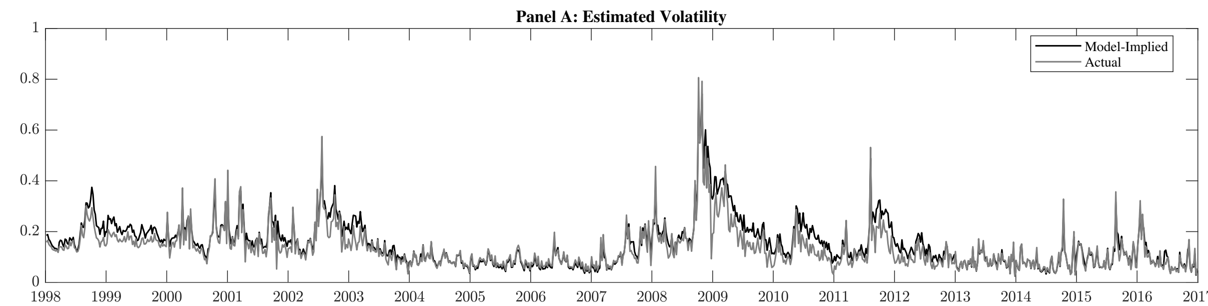
The figure shows net call (Panel A) and put (Panel B) positions of intermediaries from 1990 to 2021.

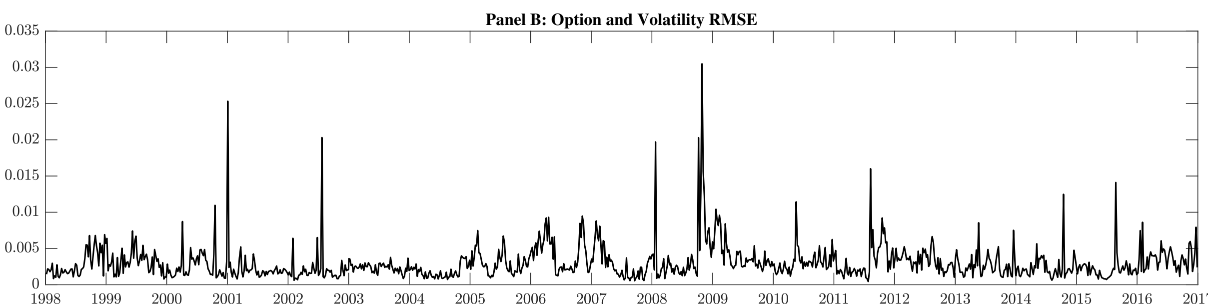
Figure G.3. Time-Series of BSM Hedge Error



The figure shows 30-day moving averages of relative pricing errors using the Black and Scholes (1973) given the current volatility surface. The error is calculated as the median of the relative difference between observable option prices and the price implied by the model.

Figure G.4. In-Sample Model Fit





The figure shows the estimated volatility defined as $\sqrt{V_{1,t} + V_{2,t}}$ compared to realized volatility estimated from high-frequency data in Panel A. The total RMSE from loss function (37) is plotted in Panel B.

Table G.2. Option Intermediary Risk and Volatility

	CV_{t+1}		VIX_{t+1}		VRP_{t+1}		LTV_{t+1}	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
α	2.81***	2.54***	2.85***	2.78***	1.28***	0.94	1.20***	1.01**
	(0.11)	(0.61)	(0.11)	(0.64)	(0.10)	(0.58)	(0.11)	(0.49)
OIR	0.35***	0.34***	0.42***	0.41***	0.36***	0.34***	0.32***	0.31***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
$Vega_{t+1}$		0.03		0.01		0.03		0.02
		(0.06)		(0.06)		(0.06)		(0.05)
r_{t+1}^{SP500}		−0.04**		−0.14***		−0.13***		−0.09**
		(0.02)		(0.02)		(0.02)		(0.03)
ΔY_{t+1}		0.09***		0.00		0.17***		0.15***
		(0.01)		(0.01)		(0.02)		(0.02)
adj. R^2	12.3	13.6	17.3	19.2	12.8	19.9	10.2	13.5
N	7511	7511	7532	7532	7511	7511	5820	5820

The table reports results of daily predictive OLS regression of risk proxies on option intermediary risk OIR. The 30-day conditional volatility expectation CV is from Bekaert and Hoerova (2014), VIX is the CBOE's volatility index, VRP is the difference between both, and LTV is left-tail variation measure from Bollerslev, Todorov, and Xu (2015). The control variables are vega-weighted inventory risk (Fournier and Jacobs, 2020), the log-return of the S&P 500, and the change of the respective dependent variable ΔY. Standard errors (in parentheses) are computed based on the method of Newey and West (1987) with 30 lags. ***, **, and * denote significance at 1%, 5%, and 10%. The sample period is from Jan 1991 until Dec 2020, except for all regressions including LTV (May 1996). All variables are standardized by their sample standard deviation.

Table G.3. Robustnesss: End-of-Day Option Returns

	Ref. Point		Level		IV Scenario		Low Vola		Realized P&L	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
OIR	-1.20^{***}	-0.53^{**}	-1.61^{***}	-0.48^{**}	-2.60^{***}	-0.53^{**}	-1.38^{***}	-0.56^{***}	-1.38^{***}	-0.50^{**}
	(0.44)	(0.23)	(0.53)	(0.21)	(0.81)	(0.23)	(0.50)	(0.22)	(0.50)	(0.22)
VIX	-1.96^{**}	3.63^{***}	-1.91^{**}	3.59^{***}	-1.67^{*}	3.63^{***}	-5.59^{***}	3.25^{**}	-1.11	3.35^{***}
	(0.94)	(1.00)	(0.93)	(0.99)	(0.89)	(1.00)	(2.11)	(1.56)	(0.81)	(0.93)
$Vega$	1.04^{*}	-2.03^{***}	1.23^{*}	-2.09^{***}	1.38^{*}	-2.02^{***}	1.10^{**}	-1.21^{**}	1.10^{*}	-1.56^{***}
	(0.61)	(0.59)	(0.65)	(0.60)	(0.70)	(0.59)	(0.53)	(0.51)	(0.65)	(0.51)
VRP	-0.05	-1.58^{**}	-0.07	-1.56^{**}	-0.12	-1.59^{**}	0.75	1.10	-0.40	-1.44^{**}
	(0.61)	(0.69)	(0.61)	(0.68)	(0.60)	(0.69)	(1.69)	(1.82)	(0.56)	(0.66)
$Skew$	-0.05	0.17	-0.06	0.14	0.05	0.20	0.71	0.02	-0.26	0.45
	(0.58)	(0.57)	(0.58)	(0.56)	(0.58)	(0.57)	(0.47)	(0.47)	(0.58)	(0.57)
r_t^{SP500}	-3.25^{**}	-3.37^{**}	-3.25^{**}	-3.37^{**}	-3.25^{**}	-3.37^{**}	-4.01	-4.00	-3.11^{**}	-3.22^{**}
	(1.28)	(1.31)	(1.28)	(1.31)	(1.28)	(1.31)	(2.74)	(2.78)	(1.26)	(1.30)
OI_t	0.27^{**}		0.27^{**}		0.30^{***}		0.40^{***}		0.24^{**}	
	(0.10)		(0.11)		(0.11)		(0.11)		(0.10)	
$P\&L_t$									-2.61^{***}	-2.48^{***}
									(0.56)	(0.54)
adj. R^2	9.12	9.27	8.96	9.27	9.02	9.26	10.18	10.08	9.62	9.85
adj. R^2 w/o	3.47	3.74	3.48	3.74	3.47	3.74	2.49	2.31	4.23	4.42
Identification	Lag	Inst	Lag	Inst	Lag	Inst	Lag	Inst	Lag	Inst
Entity FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	132,460	132,460	132,460	132,460	132,460	132,460	92,610	92,610	132,259	132,259

The table reports results of panel regression of end-of-day option returns on option intermediary risk OIR. The robustness Ref. Point uses OIR determined at the scenario of -5%, Level uses the P&L level instead of slope, IV Scenario uses OIR determined with the dynamic IV scenario, Low Vola excludes VIX days above 20%, and Realized P&L controls for the realized profit and loss of intermediaries' delta-hedged option positions. Identification Lag indicates that OIR and the other variables without a time-index are lagged by one day. Identification Inst use the instrument from Equation (17). R^2 w/o is the explained variation without fixed effects. Standard errors (in parentheses) are clustered by entity and time. ***, **, and * denote significance at 1%, 5%, and 10%. The sample period is from Jan 2004 until Sep 2020. All independent variables are standardized and multiplied by 100. The top and bottom 0.5% outliers of all sets of returns are removed. The regression constant is not reported.

Table G.4. Robustness: Liquidity Compensation

	Ref. Point		Level		IV Scenario		Low Vola		Realized P&L	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
OIR	-0.16^{***}	-0.05^{***}	-0.15^{***}	-0.02^{**}	-0.07^{***}	-0.03^{***}	-0.12^{***}	0.02	-0.11^{***}	-0.02^{**}
	(0.05)	(0.01)	(0.05)	(0.01)	(0.03)	(0.01)	(0.04)	(0.01)	(0.04)	(0.01)
VIX_t	-0.03	-0.08^{**}	-0.02	-0.09^{**}	-0.05^{*}	-0.09^{**}	0.21^{*}	-0.07	-0.02	-0.09^{**}
	(0.03)	(0.04)	(0.02)	(0.04)	(0.03)	(0.04)	(0.12)	(0.11)	(0.02)	(0.04)
SI_t	0.02^{*}	-0.01	0.02^{*}	-0.01	0.00	-0.01	-0.13	-0.11	0.00	-0.01
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.09)	(0.08)	(0.01)	(0.01)
FI_t	0.14^{***}	0.20^{***}	0.14^{***}	0.20^{***}	0.18^{***}	0.20^{***}	0.37^{***}	0.42^{***}	0.16^{***}	0.20^{***}
	(0.05)	(0.06)	(0.05)	(0.06)	(0.06)	(0.06)	(0.10)	(0.12)	(0.05)	(0.06)
$ES_{i,t-1}$	0.44^{***}	0.47^{***}	0.44^{***}	0.47^{***}	0.46^{***}	0.47^{***}	0.36^{***}	0.38^{***}	0.45^{***}	0.47^{***}
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.08)	(0.08)	(0.06)	(0.06)
$ES_{i,t-2}$	0.39^{***}	0.41^{***}	0.39^{***}	0.41^{***}	0.41^{***}	0.41^{***}	0.34^{***}	0.35^{***}	0.40^{***}	0.41^{***}
	(0.04)	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
$Volume_t$	-0.24^{***}	-0.25^{***}	-0.25^{***}	-0.26^{***}	-0.27^{***}	-0.26^{***}	-0.37^{***}	-0.37^{***}	-0.26^{***}	-0.26^{***}
	(0.08)	(0.09)	(0.08)	(0.09)	(0.09)	(0.09)	(0.11)	(0.12)	(0.09)	(0.09)
$P\&L_t$									0.01	0.01
									(0.03)	(0.03)
adj. R^2	31.28	30.81	31.24	30.77	30.92	30.78	32.05	32.05	31.06	30.77
adj. R^2 w/o	18.02	16.86	17.90	16.77	17.09	16.78	20.07	16.77	17.43	16.77
Identification	Lag	Inst	Lag	Inst	Lag	Inst	Lag	Inst	Lag	Inst
Entity FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	40,070	40,070	40,070	40,070	40,070	40,070	27,744	27,744	40,070	40,070

The table reports results of panel regressions of IV spreads on option intermediary risk OIR. The robustness Ref. Point uses OIR determined at the scenario of -5%, Level uses the P&L level instead of slope, IV Scenario uses OIR determined with the dynamic IV scenario, Low Vola excludes VIX days above 20%, and Realized P&L controls for the realized profit and loss of intermediaries' delta-hedged option positions. Lag indicates that OIR is lagged by one day. Inst is the instrument from Equation (17). R^2 w/o is the explained variation without fixed effects. Standard errors (in parentheses) are clustered by entity and month. ***, **, and * denote significance at 1%, 5%, and 10%. The sample is from Jan 2004 until Dec 2020. Independent variables are standardized and multiplied by 100. The regression constant is not reported.

Table G.5. Ask and Bid Option Returns

	Full Sample		Low Intermediation Risk		VIX > 20%	
	r_t^{bid}	r_t^{ask}	r_t^{bid}	r_t^{ask}	r_t^{bid}	r_t^{ask}
OIR_{t-1}	-1.701^{***}	-1.685^{***}	-1.353^{***}	-1.290^{***}	-4.677^{***}	-4.129^{***}
	(0.496)	(0.432)	(0.404)	(0.319)	(1.197)	(1.204)
VIX_{t-1}	-1.503^{**}	-1.817^{***}	-2.398^{***}	-2.832^{***}	55.115^{***}	45.627^{***}
	(0.606)	(0.531)	(0.871)	(0.769)	(10.242)	(8.209)
$Vega_{t-1}$	1.440^{**}	1.168^{***}	1.763^{***}	1.771^{***}	1.457	1.599^{**}
	(0.643)	(0.443)	(0.669)	(0.534)	(1.032)	(0.797)
VRP_{t-1}	-0.380	0.097	-0.019	0.575	-2.243^{***}	-1.543^{***}
	(0.266)	(0.169)	(0.562)	(0.400)	(0.534)	(0.362)
$Skew_{t-1}$	0.049	0.010	-0.216	-0.209	-5.374^{***}	-5.425^{***}
	(0.619)	(0.408)	(0.892)	(0.589)	(1.849)	(1.407)
r_t^{SP500}	-3.638^{**}	-2.950^{***}	-4.151^{**}	-3.409^{***}	-3.123^{**}	-2.554^{***}
	(1.718)	(0.976)	(1.969)	(1.094)	(1.493)	(0.795)
adj. R^2	7.79	9.61	8.72	10.92	8.84	11.74
adj. R^2 w/o	2.91	3.73	3.17	4.09	4.90	6.35
Identification	Lag	Lag	Lag	Lag	Lag	Lag
Entity FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
N	132,460	132,460	67,727	67,736	27,858	27,820

The table reports results of panel regression of end-of-day bid and ask option returns on option intermediary risk OIR and several control variables defined in Table IV. Standard errors (in parentheses) are clustered by entity and time. ***, **, and * denote significance at 1%, 5%, and 10%. The sample period is from Jan 2004 until Sep 2020. All independent variables are standardized and multiplied by 100. The top and bottom 0.5% outliers the returns are removed. The regression constant is not reported.

Table G.6. Compensation for Buy and Sell-Initiated Trades

	Low Intermediation Risk		VIX > 20%	
	$ES^{sell}_{i,t}$	$ES^{buy}_{i,t}$	$ES^{sell}_{i,t}$	$ES^{buy}_{i,t}$
OIR_{t-1}	-0.272^{**}	-0.279^{**}	-0.332^{***}	-0.438^{**}
	(0.125)	(0.120)	(0.125)	(0.174)
VIX_t	0.186	0.142^{***}	1.578	1.467^{**}
	(0.139)	(0.053)	(1.179)	(0.590)
SI_t	-0.108^{**}	-0.038	-0.086^{*}	-0.032^{***}
	(0.043)	(0.032)	(0.052)	(0.004)
FI_t	-0.105^{*}	-0.045	-0.132^{*}	0.006
	(0.054)	(0.048)	(0.071)	(0.030)
$ES^{mid}_{i,t-1}$	-0.017	-0.019	-0.019	-0.045
	(0.089)	(0.079)	(0.060)	(0.114)
$ES^{mid}_{i,t-2}$	-0.017	-0.072	-0.030	-0.043
	(0.057)	(0.079)	(0.081)	(0.090)
$Volume_t$	0.362^{***}	0.229^{***}	0.450^{**}	0.348^{***}
	(0.139)	(0.075)	(0.178)	(0.134)
adj. R^2	25.36	21.97	26.53	25.32
Identification	Lag	Lag	Lag	Lag
Entity FE	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes
N	6,243	6,243	2,664	2,664

The table reports results of panel regression of the effective relative implied volatility spread on option intermediary risk OIR and several control variables defined in Table III. Standard errors (in parentheses) are clustered by entity and time. ***, **, and * denote significance at 1%, 5%, and 10%. The sample period is from Jan 2004 until Dec 2020. All independent variables are standardized and multiplied by 100.