The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXXV, NO. 4 • AUGUST 2020

Option Profit and Loss Attribution and Pricing: A New Framework

PETER CARR and LIUREN WU*

ABSTRACT

This paper develops a new top-down valuation framework that links the pricing of an option investment to its daily profit and loss attribution. The framework uses the Black-Merton-Scholes option pricing formula to attribute the short-term option investment risk to variation in the underlying security price and the option's implied volatility. Taking risk-neutral expectation and demanding no dynamic arbitrage result in a pricing relation that links an option's fair implied volatility level to the underlying volatility level with corrections for the implied volatility's own expected direction of movement, its variance, and its covariance with the underlying security return.

A foolish consistency is the hobgoblin of little minds.
—Ralph Waldo Emerson

DIFFERENT MODELING FRAMEWORKS SERVE DIFFERENT purposes. A major focus of the existing option pricing literature is to derive option values that are internally consistent across all strikes and maturities. The literature specifies the full dynamics of the underlying security price, including the full dynamics of its instantaneous variance rate, and performs valuation of all options by taking risk-neutral expectations of their terminal payoffs. The full dynamics specification creates a single reference distribution of the relevant terminal random variable, which is then used to take the expectation. Under this approach, even if the assumed dynamics are wrong, the valuations on the option contracts remain consistent with one another relative to this erroneous reference.

*Peter Carr is with New York University. Liuren Wu is in the Department of Economics and Finance, Baruch College. We would like to thank Stefan Nagel (the Editor), the associate editor, and two anonymous referees. We would also like to thank Yun Bai; David Gershon; Kris Jacobs; Danling Jiang; Aaron Kim; Chun Lin; Jason Roth; Angel Serrat; Stoyan Stoyanov; seminar participants at Baruch College, Credit Suisse, RMIT, Stony Brook University, City University of New York Graduate Center, and the 2017 6th IFSID Conference on Derivatives for comments. Liuren Wu gratefully acknowledges the support of a grant from the City University of New York PSC-CUNY Research Award Program. We have read *The Journal of Finance's* disclosure policy and have no conflicts of interest to disclose.

Correspondence: Liuren Wu, Department of Economics and Finance, Baruch College, One Bernard Baruch Way, Box B10-225, New York, NY 10010; email: liuren.wu@baruch.cuny.edu.

DOI: 10.1111/jofi.12894

It is good to be consistent, but it is not good to be wrong. Unfortunately, the assumed dynamics of the underlying security price and its instantaneous volatility often deviate strongly from reality. For example, to price long-dated options, this approach needs to make projections on the underlying security price and its instantaneous volatility far into the future. The accuracy of long-dated projections is understandably low, and seemingly innocuous stationarity assumptions on the instantaneous volatility dynamics often generate much lower price variation (Giglio and Kelly (2018)) and much flatter implied volatility smiles (Carr and Wu (2003)) in long-term contracts than actually observed in the data.

In practice, as long as one does not hold the contracts to maturity, one does not necessarily need to make long-run predictions to trade long-dated contracts—An investor can hold a very long-dated contract for a very short period of time. In this case, the investor is less concerned about the terminal payoff than about the factors that drive profit and loss over the short holding period. In fact, the standard recommended practice of marking financial securities to market makes it vitally important that investors understand the magnitude and sources of daily value fluctuations, regardless of their intended holding period. The process of attributing an investment's profit and loss (P&L) on a given date to different risk exposures is commonly referred to as the P&L attribution process.

In this paper, we develop a new valuation framework that links the pricing of a security at a given point in time to its P&L attribution without directly referring to the terminal payoffs of the investment. The P&L attribution requires the specification of a risk structure for computing the investment's risk exposures and magnitudes. We take an investment in a European option as an example and perform the P&L attribution on the option contract via the explicit Black-Merton-Scholes (BMS) option pricing formula. Black and Scholes (1973) and Merton (1973) derive their option pricing formula by assuming constant-volatility geometric Brownian motion dynamics for the underlying security price. Their assumption does not match reality as return volatilities tend to vary strongly over time. Nevertheless, their pricing equation has been widely used by practitioners as a simple and intuitive representation of the option value in terms of its major risk sources, that is, variations in the underlying security price and its return volatility. In addition to the security price and contract terms (such as strike and maturity), the pricing equation takes a volatility input that can be used to match the observed option price. The volatility input that matches the observed option price is commonly referred to as the BMS option implied volatility.

A Taylor series expansion of the BMS option pricing formula attributes the option investment P&L to partial derivatives in time, the underlying security price, and the option's implied volatility. When the underlying security price and the option's implied volatility move continuously over time, expanding to the first order in time and second order in price and volatility is sufficient to bring the residual error to an order lower than the length of the short investment horizon.

Taking the risk-neutral expectation over the P&L attribution via the BMS pricing formula and demanding no dynamic arbitrage result in a simple pricing equation that relates the option's implied volatility level to the underlying instantaneous volatility as well as corrections due to the implied volatility's expected direction, variance, and covariance with the security return.

In contrast to the traditional option pricing approach, which links the values of all option contracts to a single reference dynamics specification, our new approach links the current fair value of one option contract's implied volatility to current conditional moments of log changes in the security price and the given contract's implied volatility. This subtle—but vital—shift in perspective is due to the use of the option's own implied volatility rather than the underlying security's instantaneous variance rate as a state variable. This new perspective allows one to completely localize the valuation of an option contract's implied volatility level at a given point in time to the moment conditions on the variation of this particular implied volatility at that point in time.

For comparison, one can regard the traditional option pricing framework as a bottom-up valuation approach, with the key focus being specifying the appropriate basis that centralizes the valuation of all contracts. The key strength of this centralized bottom-up approach is that cross-sectional consistency is maintained through the use of one set of reference dynamics. By contrast, our new framework is a decentralized top-down approach. By pricing an option contract based only on its own risk-neutral moment conditions, the theory does not impose cross-sectional consistency across different option contracts. The new approach is decidedly local in terms of the investment horizon it considers, and very much decentralized in terms of the contract it values.

Under this new approach, to compare the valuation of two or more distinct option contracts, one must first compare the risk exposures and magnitudes of these contracts. One can impose common factor structures on the moment conditions of different contracts to link the valuations of these contracts. Nevertheless, their commonality, or lack thereof, is left as an empirical determination or theoretical construction, but not a no-arbitrage condition. Therefore, while the traditional bottom-up approach is about the search for the omnipotent basis reference, our new top-down approach is more about accurate forecasts of the moment conditions for the particular contract. The two approaches do not directly compete, but rather complement each other. Dynamics specifications from the traditional approach can provide insights for formulating hypotheses on moment condition estimation, while empirically identified co-movement patterns in the moment conditions can provide guidance for the specification of reference dynamics.

To illustrate, we explore the cross-sectional pricing implications of the new framework under various commonality assumptions. First, we define the at-the-money option at a given maturity as the particular option whose log strike-forward ratio is equal to half of the option's total implied variance, which is the BMS risk-neutral mean of the log security return to maturity. The fair valuation of this at-the-money option does not depend on the variance and covariance of the implied volatility change, but depends only on its risk-neutral drift and the instantaneous variance level. By imposing a common risk-neutral drift on two

nearby at-the-money option contracts, we can extract the common risk-neutral drift from the term structure slope defined by these two nearby contracts. Alternatively, by assuming a common one-factor mean-reverting structure on the at-the-money implied variance, we can define an at-the-money implied variance term structure function as an exponentially weighted average of the short- and long-dated implied variances, analogous to implications from traditional stochastic volatility models (e.g., Heston (1993)).

At a fixed time to maturity, we take the at-the-money implied variance as the reference point and perform a vega hedge of other contracts using the at-the-money option. We can then represent the implied volatility skew of an option contract relative to the at-the-money implied volatility at this maturity as a function of the implied volatility's variance and covariance with the security return. Assuming that proportional movements have common variance and covariance within a certain strike range, we can extract the common variance and covariance from the implied volatility smile shape within this strike range.

Another way to look at the traditional option pricing framework is from the perspective of spanning as articulated by Bakshi and Madan (2000), who regard the characteristic function of the underlying security return as the basis that spans most derivative securities. The dynamics specification dictates the pricing of Arrow and Debreu (1954) securities, which spans the payoff of most contingent claims. The payoff spanning perspective has a strong cross-sectional focus. It is because of this focus that pricing errors from traditional option pricing models are often regarded as the starting point for statistical arbitrage trading (e.g., Duarte, Longstaff, and Yu (2007), Bali, Heidari, and Wu (2009)). By contrast, our new pricing framework focuses more on the risk-return trade-off for a particular contract, in that the pricing of the contract is made consistent with one's view on that contract's risk exposures and magnitudes, rather than with the pricing of other contracts. Because of this different perspective, our new approach can take outsourced expert forecasts on risks and risk premiums underlying a particular investment, convert them into risk-neutral moment conditions, and directly generate pricing implications without needing to understand the source or rationale of these forecasts. Conversely, we can infer moment conditions from market prices and examine the information content of these moment conditions. In empirical analysis on S&P 500 index options, we show that the risk-neutral drift extracted from the at-the-money implied volatility term structure can be used to forecast future implied volatility movements, and the variance and covariance extracted from the implied volatility smile can be combined with historical moment estimates to generate better future realized variance and covariance forecasts. We also show that these forecasts can be incorporated profitably into investment decisions in the derivative contracts. In contrast to traditional statistical arbitrage trading based on the reversion assumption on pricing errors, the time series variance and covariance forecasts allow us to identify investment opportunities with better risk-return trade-offs. Although investors can in principle make money from both approaches, our new framework provides a complementary perspective that can potentially expand profitable investment opportunities.

As the BMS pricing formula has been widely adopted in the industry as a transformation tool, P&L attribution based on the BMS pricing equation is also common (Bergomi (2016)). There also exists a valuation method in the industry based on options' BMS vega, vanna, and volga. The method is commonly referred to as the vanna-volga model (Castagna and Mercurio (2007), Wystup (2010)). The model starts by defining a reference volatility level and generating a reference value for the option based on the BMS model. The price difference of the target contract from the reference value is assumed to be linear in the price differences of three pillar options from their respective reference values. The three coefficients are determined by equating the vega, vanna, and volga of the target option and the portfolio of the three pillar options, with all of the greeks evaluated at the reference volatility. The method has been used to value both European options and exotic options based on observed prices on three appropriately chosen pillar options. It relies on the idea that the risk exposures of an option contract, defined via the BMS representation, can be approximately spanned by the risk exposures of three appropriately chosen option contracts. The vanna-volga model is similar to our approach in that it relies on the BMS pricing representation in defining risk exposures, but it is different otherwise in terms of both implementation and perspective. Our BMS risk exposures on a contract are computed against the implied volatility of that contract instead of a reference volatility, and our pricing result links the implied volatility level of an option contract to its own risk-neutral moment conditions, rather than linking the value of one option contract to the values of three other option contracts. Indeed, the spanning perspective of the vanna-volga model is much closer to the traditional option pricing approach than to our risk-return trade-off perspective.

The industry has also proposed several variations of the vanna-volga pricing approach. For example, like us, Arslan et al. (2009) and Gershon (2018) calculate the BMS risk exposures of a contract using the implied volatility of that contract instead of a reference volatility. However, both of these papers adopt the spanning perspective of the vanna-volga model and demand consistency with the observed prices of three pillars.

Also related is the strand of the practitioner literature that attempts to directly model the implied volatility dynamics for the pricing of exotic contracts.[1] These attempts, often referred to as market models of implied volatility, take the observed implied volatility as given while specifying the continuous martingale component of the volatility surface. From these two inputs, they seek to derive the no-arbitrage restrictions on the risk-neutral drift of the implied volatility dynamics. The approach is analogous to the Heath, Jarrow, and Morton's (1992) model on forward interest rates and can in principle be used to price derivatives written on the implied volatility surface. What prevents these attempts from achieving their objective is that knowledge about the shape of the current implied volatility surface places constraints on the specification of the

[1] See, for example, Avellaneda and Zhu (1998), Schönbucher (1999), Hafner (2004), Fengler (2005), and Daglish, Hull, and Suo (2007).

continuous martingale component for its future dynamics. In this paper, rather than ignoring these constraints, we fully exploit them in building a simple, direct linkage between the current level of the implied volatility and its first and second risk-neutral moment conditions.

In the academic literature, Israelov and Kelly (2017) recognize the limitations of standard option pricing models and propose directly predicting the distribution of the option investment return empirically. Earlier empirical analysis of option investment returns includes Coval and Shumway (2001) and Jones (2006). Several recent studies link option returns to various firm characteristics: An et al. (2014) link future option implied volatility variation to past stock return performance, Boyer and Vorkink (2014) link ex post option returns to ex ante implied volatility skew, Byun and Kim (2016) link option returns to the underlying stock's lottery-like characteristics, and Hu and Jacobs (2017) link option returns and the underlying stock's volatility level. Our research shares the same shift of focus from terminal payoffs to the behavior of short-term investment returns. Our theory provides a foundation for how to analyze option investment returns and how to link the predicted investment return behavior to its pricing.

The remainder of this paper is organized as follows. Section I introduces the notation and establishes the P&L attribution for an option contract based on the BMS pricing formula. Section II takes the risk-neutral expectation of the return attribution and derives the no-arbitrage pricing implications. Section III examines the cross-sectional pricing implications of the theory under different local and global commonality assumptions. Section IV performs empirical analysis on S&P 500 index options and explores different applications of the new theory. Section V discusses the theory's general implications, its limitations, and the risk representations. Section VI concludes.

I. P&L Attribution on Option Investments

We consider a market with a risk-free bond, a risky asset, and one vanilla European option written on the risky asset. For simplicity, we assume zero interest rates and zero carrying costs/benefits on the risky asset. In practical implementation, one can readily accommodate a deterministic term structure of financing rates by modeling the forward value of the underlying security and defining moneyness of the option against the forward. The risky asset can be any type of tradable security, but for concreteness we refer to it as the stock. In the United States, exchange-traded options on individual stocks are American style. To apply our new theory to American options, a commonly used shortcut is to extract the BMS implied volatility from the price of an American option based on some tree/lattice method and use the implied volatility to compute a European option value for the same maturity date and strike.[2]

[2] See Carr and Wu (2010) for a detailed discussion on data preprocessing of individual stock options.

We assume frictionless and continuous trading in the risk-free bond, the stock, and the option contract written on the stock. We assume no-arbitrage between the stock and the bond. As a result, there exists a risk-neutral probability measure $\mathbb{Q}$, equivalent to the statistical probability measure $\mathbb{P}$, such that the stock price S is a martingale. We further assume that the option value we seek does not allow arbitrage against any portfolio of the stock and the riskless bond.

We start by considering a long position in a European call option. Holding the call to expiry generates a P&L dictated by the terminal payoff of the call. Classic option valuation starts with the terminal payoff function and takes the expectation of the terminal payoff based on assumptions governing the dynamics of the underlying asset price. Our new pricing approach focuses on the instantaneous investment P&L over the next instant. The short-term P&L fluctuation of the option investment is driven mainly by the option's exposures to various risk sources and the variation in these risk sources. Accordingly, our analysis focuses more on defining risk exposures and quantifying risk magnitudes than on describing terminal payoffs.

We attribute the short-term investment P&L on the option contract by making use of the explicit BMS pricing formula. The formula attributes the variation in an option contract's value to variation in the calendar time t, the underlying security price S_t, and the option's BMS implied volatility I_t. Formally, let $B(t, S_t, I_t; K, T)$ denote the BMS representation of the option value as a function of the three variates (t, S_t, I_t) for a European call option contract with strike price K and expiry date T. The pricing formula is given by

$$B(t, S_t, I_t; K, T) \equiv S_t N\left(-\frac{k - \frac{1}{2} I_t^2 \tau}{I_t \sqrt{\tau}}\right) - K N\left(-\frac{k + \frac{1}{2} I_t^2 \tau}{I_t \sqrt{\tau}}\right), \tag{1}$$

where $N(\cdot)$ denotes the cumulative normal function, $\tau \equiv T - t$ the time to maturity, and $k \equiv \ln(K/S_t)$ the relative strike. We henceforth use the terms $z_\pm \equiv (k \pm \frac{1}{2} I_t^2 \tau)$ to represent the convexity-adjusted moneyness of the call under the risk-neutral measure and the share measure, respectively, in the BMS model environment.

For a given option contract, the BMS pricing formula represents the option value at time t as a function of the stock price S_t and the option implied volatility I_t. As long as the option price does not allow arbitrage against the underlying risky stock and the riskless bond, one can always find a positive implied volatility input to the BMS pricing formula to match that price (Hodges (1996)). The BMS pricing formula builds a monotonic linkage between the option price and the option's implied volatility, and captures all random shocks to the option (other than shocks to the underlying security price level) through the implied volatility.

Using the BMS pricing equation, we can attribute the instantaneous P&L of the option investment to the variation in the calendar time, the stock price,

and the implied volatility,

$$\mathrm{d}B = \left[B_t \mathrm{d}t + B_S \mathrm{d}S_t + B_I \mathrm{d}I_t\right] + \left[\frac{1}{2}B_{SS}(\mathrm{d}S_t)^2 + \frac{1}{2}B_{II}(\mathrm{d}I_t)^2 + B_{IS}(\mathrm{d}S_t \mathrm{d}I_t)\right] + J_t, \tag{2}$$

where we suppress the arguments $(t, S_t, I_t; K, T)$ of the pricing function and use B_t, B_S, B_I, B_{SS}, B_{II}, and B_{IS} to denote the partial derivatives (sensitivities) of the pricing function against the calendar time t, the security price S, and the option's implied volatility I. The partial derivatives are commonly referred to as the option's theta (B_t), delta (B_S), vega (B_I), gamma (B_{SS}), volga (B_{II}), and vanna (B_{IS}), respectively. The first bracket collects first-order derivatives, the second bracket collects second-order derivatives, and the last term J_t captures the contribution of potential higher order derivatives due to random jumps in the stock price and option implied volatility. When both the stock price and the option implied volatility show purely continuous movements, the first and second derivatives capture all of the relevant movements for the option price over a short time interval. We henceforth assume continuous dynamics and link option pricing to its first- and second-order derivatives/exposures.

II. Risk-Neutral Expectation and Implied Volatility Valuation

We assume continuous dynamics, take the expectation of the option P&L attribution in equation (2) under the risk-neutral measure $\mathbb{Q}$, and divide the expected P&L by the instantaneous investment horizon $\mathrm{d}t$,

$$\frac{\mathbb{E}_t[\mathrm{d}B]}{\mathrm{d}t} = B_t + B_I I_t \mu_t + \frac{1}{2}B_{SS}S_t^2\sigma_t^2 + \frac{1}{2}B_{II}I_t^2\omega_t^2 + B_{IS}I_t S_t \gamma_t, \tag{3}$$

where $\mathbb{E}_t[\cdot]$ denotes the expectation operation under the risk-neutral measure conditional on time-t filtration and μ_t the annualized risk-neutral expected rate of change in the BMS implied volatility of the option contract,

$$\mu_t \equiv \mathbb{E}_t\left[\frac{\mathrm{d}I_t}{I_t}\right]/\mathrm{d}t, \tag{4}$$

and σ_t^2, ω_t^2, and γ_t the time-t conditional variance and covariance rate of the stock return and the implied volatility change,

$$\sigma_t^2 \equiv \mathbb{E}_t\left[\left(\frac{\mathrm{d}S_t}{S_t}\right)^2\right]/\mathrm{d}t, \quad \omega_t^2 \equiv \mathbb{E}_t\left[\left(\frac{\mathrm{d}I_t}{I_t}\right)^2\right]/\mathrm{d}t, \quad \gamma_t \equiv \mathbb{E}_t\left[\left(\frac{\mathrm{d}S_t}{S_t}, \frac{\mathrm{d}I_t}{I_t}\right)\right]/\mathrm{d}t. \tag{5}$$

Under the zero financing cost assumption, the risk-neutral expected stock return is zero.

The zero financing cost assumption and no dynamic arbitrage dictate that the risk-neutral expected return on the option investment is also zero,

$$0 = B_t + B_I I_t \mu_t + \frac{1}{2} B_{SS} S_t^2 \sigma_t^2 + \frac{1}{2} B_{II} I_t^2 \omega_t^2 + B_{IS} I_t S_t \gamma_t. \tag{6}$$

We can regard equation (6) as a pricing relation. As long as the option price satisfies no dynamic arbitrage and generates a risk-neutral expected return of zero, the option price must satisfy the constraints imposed by this equation. This pricing equation is not based on the full specification of the underlying security price dynamics, but rather on the first and second conditional moments of the security price and the option's implied volatility movements at time t.

The pricing relation in (6) highlights the short-term trade-offs among the different sources of expected gains and losses from holding an option. By being long in an option, one loses time value as calendar time passes. The rate of loss is captured by the option's theta, B_t. This theta loss is compensated by expected gains from the security price variation, measured by the security's return variance σ_t^2, due to the option's positive gamma exposure B_{SS}. Variation in the option's implied volatility (ω_t^2) and in its covariation with the security return (γ_t) induce additional expected gains or losses due to the option's volga (B_{II}) and vanna (B_{IS}) exposures, respectively. The option's positive vega exposure (B_I) can be another source of expected gain or loss depending on the expected direction (μ_t) of the implied volatility movement. The no-arbitrage condition dictates that the option be priced such that at any point in time these different sources of expected gains and losses balance out to produce a zero risk-neutral expected excess return on the investment.

THEOREM 1: *Under continuous price and implied volatility movements and zero financing costs, no dynamic arbitrage requires that an option be priced to balance out the option's theta loss against the expected gains and losses from the option's vega, gamma, volga, and vanna exposures at any point in time,*

$$-B_t = B_I I_t \mu_t + \frac{1}{2} B_{SS} S_t^2 \sigma_t^2 + \frac{1}{2} B_{II} I_t^2 \omega_t^2 + B_{IS} I_t S_t \gamma_t. \tag{7}$$

The gains and losses from these exposures are determined by the expected rate of the option's implied volatility movement (μ_t), the security's return variance σ_t^2, the implied volatility's variance ω_t^2, and the covariance between the security return and the implied volatility γ_t, respectively.

Under this pricing relation, determining whether an option is fairly priced at any point in time amounts to determining whether the option's exposures are balanced out with the associated first- and second-moment forecasts at that time. The pricing relation does not specify how to determine these forecasts or how the forecasts vary over time. Thus, an interesting feature of the pricing relation is that the risk forecasting process can be completely separated from the pricing process.

COROLLARY 1: *Under continuous price movements and constant implied volatility for an option contract with zero financing, the theta loss is compensated fully by the gamma gain,*

$$-B_t = \frac{1}{2} B_{SS} S_t^2 \sigma_t^2.$$

When the option's implied volatility does not move over time, the only source of variation comes from the security price and hence the theta loss is balanced out completely by the gamma gain. This is the case under the BMS model environment when the instantaneous return volatility σ is a constant.

A particularly nice feature of the BMS pricing equation is that the BMS theta (B_t), cash vega ($B_I I_t$), cash vanna ($B_{IS} I_t S_t$), and cash volga ($B_{II} I_t^2$) can all be represented in terms of the BMS cash gamma ($B_{SS} S_t^2$),

$$\begin{aligned} B_t &= -\frac{1}{2} I_t^2 B_{SS} S_t^2, \qquad B_I I_t = I_t^2 \tau B_{SS} S_t^2, \\ B_{IS} I_t S_t &= z_+ B_{SS} S_t^2, \quad B_{II} I_t^2 = z_+ z_- B_{SS} S_t^2. \end{aligned} \tag{8}$$

The Appendix provides the derivation. For an option contract with strictly positive cash gamma, we can define the option investment return as the investment P&L per unit of cash gamma so that we can factor out the cash gamma component from the pricing relation in (7) to obtain the following algebraic equation

$$I_t^2 = \left[2\tau \mu_t I_t^2 + \sigma_t^2\right] + \left[2\gamma_t z_+ + \omega_t^2 z_+ z_-\right]. \tag{9}$$

THEOREM 2: *Assuming continuous price and implied volatility movements, and performing instantaneous P&L attribution on a European option investment based on the BMS pricing equation, we arrive at a no-arbitrage pricing relation between the time-t fair value of the option's implied volatility level and the time-t risk-neutral conditional mean and variance of the implied volatility percentage change (μ_t and ω_t^2), the conditional variance of the underlying security return (σ_t^2), and the conditional covariance between the two (γ_t).*

Compared to traditional option pricing practice, equation (9) represents an extremely simple formulation of the option's fair value. Traditional option pricing emphasizes cross-sectional consistency as its chief objective. To achieve this objective, one specifies the full risk-neutral dynamics on the underlying security and prices all options by taking expectations of their terminal payoff functions based on the same dynamics assumption. These assumed dynamics on the underlying security serve as a single yardstick for all option contracts. The same yardstick guarantees that valuations of all options are consistent with this yardstick and hence with one another.

By contrast, by starting with a localized P&L attribution of one option contract, equation (9) only guarantees that the time-t valuation of this contract is consistent with the first and second risk-neutral conditional moments of the underlying security price and this option's implied volatility. It guarantees no

dynamic arbitrage between this option contract and the underlying security and cash under the assumed moment conditions, but nothing more. Thus, instead of providing a yardstick for analyzing cross-sectional consistency across different option contracts, the new approach provides a more direct, top-down linkage between the fair implied volatility valuation of an option contract at any given point in time and its conditional moment conditions at that time. In this sense, while traditional option pricing focuses on cross-sectional consistency across contracts, the new pricing relation focuses on the risk-return trade-off for one particular contract.

III. Commonality and Cross-Sectional Pricing

The pricing relation in (9) applies to one particular option contract, with no reference to any other option contracts. This relation allows one to compare the valuation of an option contract to one's projection of the first- and second-moment conditions of the underlying security price and the option's implied volatility. To make valuation comparisons across different option contracts under this framework, one must first make comparisons on the first and second conditional moments of the corresponding option implied volatilities.

Under the BMS model environment, the implied volatilities of all options underlying the same security are the same. In reality, option implied volatilities at different strikes and maturities can differ; nevertheless, they tend to move together. Furthermore, the implied volatility levels tend to be closer to each other and their co-movements tend to be stronger as the strike and maturity distances between the option contracts become smaller. As such, the implied volatility surface across strikes and maturities tends to possess a smooth shape. These basic observations can serve as a starting point in analyzing the commonalities and differences in the implied volatility movements across strikes and maturities underlying a given security.

In this section, under our new pricing framework, we examine the cross-sectional pricing implications of different commonality assumptions. We examine the validity of some of the commonality assumptions in Section IV. The analysis demonstrates how the new pricing framework generates cross-sectional implications based on explicit commonality assumptions on the underlying moment conditions.

A. The At-the-Money Implied Variance Term Structure

To separate the term structure effect from the moneyness effect, we define the at-the-money option as the option with $z_+ = k + \frac{1}{2} I_t^2 \tau = 0$, which corresponds to the strike price that equates the relative strike k to the risk-neutral expected value of $\ln(S_T/S_t)$ under the BMS model environment.

At $z_+ = 0$, equation (8) shows that the option has both zero volga and zero vanna. Investing in such an option only exposes the investor to delta, vega, and gamma risk. The pricing equation in equation (9) also simplifies as the terms in the second bracket are reduced to zero. If we use a different notation, A_t, to

denote the at-the-money implied volatility at a certain time to maturity τ,[3] we can write the pricing equation for the at-the-money implied volatility as

$$A_t^2 = 2\tau\mu_t A_t^2 + \sigma_t^2. \tag{10}$$

The at-the-money option with $z_+ = 0$ is the only strike point at which the option has both zero volga and zero vanna. It is also the only strike point at which the implied volatility level depends only on the risk-neutral expected rate of change of the implied volatility, but not on its variance and covariance with the security return. The separation allows us to analyze expected volatility changes and the term structure without interference from the second-order effects.

To extract the expected rate of change from the at-the-money implied volatility term structure, we propose to make the following local commonality assumption.

ASSUMPTION (Local commonality on rates of change): The expected rates of change for at-the-money implied volatilities of nearby maturities are the same,

$$\mu_t(\tau_1) \doteq \mu_t(\tau_2), \tag{11}$$

when $|\tau_1 - \tau_2|$ is small.

With the local commonality assumption, we can extract the common expected rate of change from the implied variance slope within this maturity range.

PROPOSITION 1: *When the implied volatilities of at-the-money option contracts within a maturity range* $[\tau_1, \tau_2]$ *share the same risk-neutral expected rate of change* μ_t *at time t, this common rate can be extracted from the at-the-money implied variance slope within this maturity range,*

$$\mu_t = \frac{A_t^2(\tau_2) - A_t^2(\tau_1)}{2\left(A_t^2(\tau_2)\tau_2 - A_t^2(\tau_1)\tau_1\right)}. \tag{12}$$

Equation (12) can be readily derived by applying equation (10) twice at the two maturities τ_1 and τ_2 with the common expected rate of change μ_t.

The accuracy and validity of the extracted common rate of change depends on both the data quality of the implied volatility observations and the actual stability of the rate of change across the term structure. In representing the local commonality assumption, equation (11) uses the approximate equal sign "$\doteq$" to highlight the approximate nature of the assumption. Similarly, the term "local" is not an exact phrasing, either. The extracted rate of change from (12) represents an approximate estimate of the true underlying rate of change, with the approximation error decreasing as the maturity distance decreases and when the variation in the rate of change is smaller. In Section IV, we

[3] Where possible confusion is not a concern, we drop the functional dependence of A_t on the time to maturity to reduce notational clutter.

examine when the local commonality assumption holds reasonably well and when the assumption can break down.

B. The Implied Volatility Smile

To isolate the moneyness effect from the term structure effect, we consider vega hedging with the at-the-money contract of the same expiry, assuming strong implied volatility co-movements for options at the same expiry. In doing so, we take the at-the-money implied volatility level A_t as given, and represent the implied volatility levels of other option contracts at the same expiry as spreads (skews) relative to the at-the-money implied volatility level.

In particular, if we assume that the expected rate of change scales proportionally with the at-the-money contract according to

$$\mu_t I_t^2 = \mu_t^A A_t^2, \tag{13}$$

we can subtract (10) from (9) to highlight the implied volatility smile effect across moneyness,

$$I_t^2 - A_t^2 = 2\gamma_t z_+ + \omega_t^2 z_+ z_-. \tag{14}$$

At each strike and accordingly each moneyness level $z_\pm$, the variance and covariance of the implied volatility change (ω_t^2, γ_t) jointly determine the degree to which the implied variance level at that strike I_t^2 deviates from the at-the-money implied variance level A_t^2. Thus, the shape of the implied volatility smile at any point in time is driven by how much the implied volatility is expected to move over the next instant and how it is expected to co-move with the security price. If one expects the implied volatility to move a lot in the next instant, one cannot expect the implied volatility smile to be flat.

As the variance and covariance (ω_t^2, γ_t) jointly determine the shape of the implied volatility smile, we can reverse-engineer and extract the market-expected variance and covariance from the observed shape of the implied volatility smile. To do so, we make local commonality assumptions on the movements of the implied volatilities across moneyness.

ASSUMPTION (Local commonality on variance and covariance rates): The variance and covariance rates of implied volatilities across a range of strikes at the same maturity are the same,

$$\omega_t^2(k) \doteq \omega_t^2, \gamma_t^2(k) \doteq \gamma_t, \tag{15}$$

for all k within a certain strike range.

As near-the-money options are the most actively traded, it is of high practical interest to extract the variance and covariance of the at-the-money implied volatility by assuming common variance and covariance for strikes within a certain range of the forward, or equivalently, for the convexity-adjusted moneyness measure z_+ to be within a certain range around zero.

It is often observed that the implied volatility smile tends to have a smooth shape, and that the implied volatilities at nearby strikes tend to co-move strongly, a stylized observation that we document in Section IV. The local commonality assumption in (15) amounts to assuming that the implied volatilities at nearby strikes move by the same expected proportional magnitude.

PROPOSITION 2: *With the local commonality assumption in (13) and (15), we can estimate the locally common variance and covariance from a simple local linear cross-sectional regression of the implied variance spread* $(I_t^2 - A_t^2)$ *on the two convexity-adjusted moneyness measures* $[2z_+, z_+z_-]$.

In Section IV, we estimate the variance and covariance of the at-the-money implied volatility from a cross-sectional regression within a narrow range of moneyness around zero, and we examine their information content in predicting future variation in implied volatility.

IV. Empirical Analysis on S&P 500 Index Options

We perform empirical analysis on S&P 500 index (SPX) options. SPX options are actively traded on the Chicago Board of Options Exchange (CBOE). We obtain the history of closing option prices and implied volatilities on SPX options, as well as the underlying index level and interest rate series, from the data vendor OptionMetrics. The sample period is from January 4, 1996 to April 29, 2016, which spans 5,111 business days. Over the sample period, the index level started at around 617 and ended at around 2,065, with an annualized daily return volatility estimate of 19.5%.

The new theory prices an option on the basis of the first and second risk-neutral moment conditions of the implied volatility changes of the particular contract. As the data comprise quotes on exchange-traded option contracts, we can easily compute the implied volatility changes of an option contract following its variation over consecutive days. To examine how the statistical moment conditions of the implied volatility changes vary across moneyness and time to maturity, and how such variation is related to the pricing of the implied volatility term structure and the implied volatility smile, we construct floating series on the implied volatility changes at different moneyness and time-to-maturity grids via local smoothing interpolation. First, we provide evidence on local commonality in the co-movements among the floating implied volatility change series. Second, based on the local commonality observations, we extract the locally common expected rate of change from the at-the-money implied volatility term structure of adjacent maturities. We examine the information content of the extracted rate of change in predicting future implied volatility changes. Third, we construct variance and covariance estimates of the floating implied volatility time series and compare them with the locally common risk-neutral variance and covariance estimates extracted from the local implied variance smile around at-the-money. We then examine the extent to which the future realized variance and covariance estimates can be predicted by their corresponding historical time series estimates and the cross-sectionally

extracted risk-neutral estimates. We also construct a trading strategy based on the difference between the future variance and covariance forecasts and the risk-neutral moments currently priced in the local smile. We show that the trading strategy can capture large risk premiums with high information ratios, and that this strategy is distinct from traditional statistical arbitrage strategies.

A. Construct Floating Series of Implied Volatility Percentage Changes

The new theory relates the fair level of the implied volatility of an option contract to the first and second risk-neutral moments of the changes in the implied volatility of that contract. As the data we have are quotes on fixed option contracts, we can readily compute their implied volatility changes by following the pricing of each contract over consecutive days. Nevertheless, to examine how the moment conditions of these implied volatility changes vary across moneyness and time to maturity, we need to construct floating time series on implied volatility changes at fixed moneyness and time-to-maturity grids. The constructed floating series also allow us to better explore both the global factor structures and the local commonalities in the implied volatility changes across the moneyness and time-to-maturity grids.

Most prior empirical studies examine the behavior of floating implied volatility time series at fixed time to maturity and fixed moneyness. Changes in such floating implied volatility time series can be quite different from the fixed-contract implied volatility changes that we construct. The difference can be particularly large when the implied volatility surface has a steep term structure and/or a strong implied volatility skew or smile. Even if the implied volatility of each option contract were to remain the same, sliding along the term structure (due to time running forward) and along the skew (due to spot price movement) would still lead to large changes in the floating implied volatility series. Conversely, even if the implied volatility surface—as a function of time to maturity and standardized moneyness—were to remain the same, the implied volatility for a particular option contract with fixed expiry date and strike price can change as its time to maturity and its relative moneyness vary over time.

From the observations on fixed option contracts, we construct floating time series on both the implied volatility levels and the fixed-contract implied volatility changes via local smoothing interpolation. Specifically, at each date t, for each option contract i available on that date, we retrieve its implied volatility on both date t, I_t^i, and the next business date $t+1$, I_{t+1}^i, and we compute the log percentage implied volatility change, $R_{t+1}^i \equiv \ln(I_{t+1}^i / I_t^i)$, on this option contract. We then smooth interpolate to generate both the time-t level and the change at floating time to maturity and moneyness points.

We construct the floating time series at a grid of five time-to-maturity levels and nine moneyness levels. The chosen time-to-maturity (τ) grids are at one month (30 days), two months (60 days), three months (91 days), six months

(182 days), and 12 months (365 days). The maturity spacing choice is similar to common industry practice with finer grid points at shorter maturities where option trading activity is concentrated and where the term structure varies the most.

At each maturity, the moneyness grids are constructed based on the standardized moneyness measure $x \equiv z_+/I_t\sqrt{\tau}$, which can be interpreted as the number of standard deviations by which the log strike $\ln(K)$ exceeds the mean of the log terminal price $\ln(S_T)$ under the BMS model environment. We build the grids up to two standard deviations with a uniform interval of half a standard deviation, $\mathbf{x} = 0, \pm 0.5, \pm 1, \pm 1.5, \pm 2$.

We estimate both the implied volatility level I_t and the log implied volatility change R_{t+1} of an option contract at each maturity-moneyness grid (τ, x) via local averaging with the following weighting schemes. First, at each strike, there can be two quotes, one from the call option and the other from the put option. We put more weight on the out-of-the-money option contract, which tends to be more actively traded and hence tends to have a more reliable quote. We use one minus the absolute value of the option's forward delta as the weight, and further truncate the weight to zero when the absolute delta is greater than 80%. The truncation sets the weights on deep in-the-money options to zero when the absolute delta is over 80%, where the quotes tend to become unreliable.

Second, we weigh each observation based on its distance to the target log time to maturity $(\ln \tau)$ and its distance to the target moneyness level (x) based on an independent bivariate Gaussian kernel with default bandwidth choices. Taking logarithm on time to maturity gives more resolution to the shorter time to maturity.

Taken together, to construct the implied volatility level and percentage change of an option contract at a target maturity-moneyness grid (τ, x), we weigh each contract i according to

$$w_i = (1 - |\delta_i|) I_{|\delta_i|<.8} \exp\left(-\frac{(x_i - x)^2}{2h_x^2}\right) \exp\left(-\frac{(\ln \tau_i - \ln \tau)^2}{2h_\tau^2}\right), \tag{16}$$

where δ_i denotes the BMS forward delta of the option and (h_x, h_τ) denote the two bandwidths.

Our interpolation scheme is based on industry best practice. Common variations in the interpolation scheme include the relative weighting between call and put options at the same strike and the degree of smoothing applied to the quotes. Small variations in the interpolation methodology do not affect the general conclusion of the analysis.

B. Local Commonality and Global Factor Structures

Our new theory links an option's implied volatility level to its own first and second risk-neutral conditional moments. To extract the moment conditions from the observed implied volatility level via reverse-engineering, we propose

the concept of *local commonality*: Implied volatilities of nearby contracts (with strikes and maturities close to one another) tend to move closely together and thus share similar moment conditions. The concept is not meant to be exact as the terms "local" and "nearby" are relative and approximate concepts, very different from the exact specification of the number of factors in a traditional stochastic volatility model. Instead, it is meant to capture a common and robust qualitative feature of the implied volatility surface predicted by most stochastic volatility models. Naturally, nearby option contracts have similar option payoffs and, accordingly, similar loading coefficients on stochastic volatility factors, unless the models have peculiar calendar-day or price-threshold dependent behaviors.[4]

The implied volatilities of option contracts with large distances in maturity and strike (e.g., one-month option versus one-year option) can move either together or separately, depending crucially on the underlying factor structure of the volatility dynamics. However, implied volatilities of nearby option contracts (say, 11- and 12-month at-the-money options) tend to co-move strongly regardless of the particular specification of the underlying volatility dynamics. In the limit, the implied volatilities of option contracts with the same strike and expiry should be identical by arbitrage regardless of dynamics assumptions. The local commonality concept is an extension of this logic. It depends less on model assumption but more on contract similarity.

The concept of local commonality also captures the natural response of shrewd options market makers who tend to hedge their exposures to a derivative contract with nearby contracts. Wu and Zhu (2016) show the theoretical and empirical robustness of hedging with nearby contracts. Our local commonality assumption strives to capture this practical robustness.

Notwithstanding the natural appeal, it is important to assess how far the commonality assumption can be pushed in practice in terms of the distance in maturity and moneyness. For this purpose, we examine how co-movements of implied volatility changes across contracts vary with the distance between the contract's maturity and moneyness.

First, we measure the cross-correlation of the percentage implied volatility change series and examine how the correlation estimates vary with the distance in maturity and moneyness. To help visualize the results, Figure 1 takes the three-month at-the-money option at the center of the maturity-moneyness grid as the reference point and shows how the correlation estimates of the implied volatility changes between this contract and all other contracts vary across moneyness at different maturities. The solid line plots the correlation estimates with contracts at the same maturity. The estimates are well over 90% for contracts within one standard deviation in moneyness ($|x| \leq 1$) and

[4] An earnings event with a known announcement date represents one such calendar-day effect, which can lead to rather different implied volatility behaviors for options maturing before and after the event date, even if the maturity difference is merely one day apart. A near-term announced or rumored corporate event with a specific triggering stock price level can also potentially cause discontinuity in implied volatility behavior around the triggering strike price.

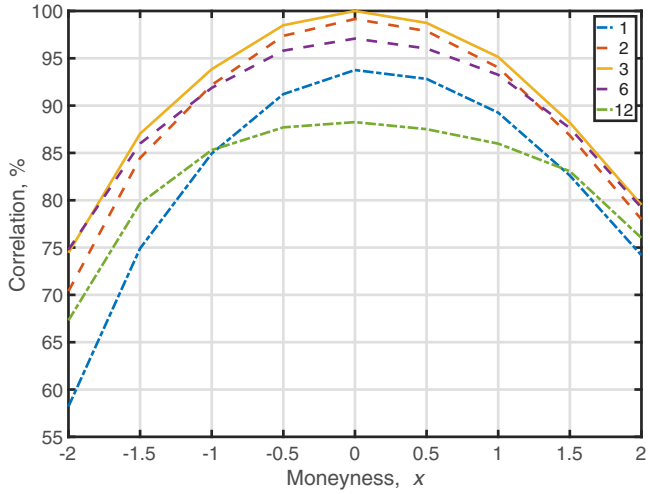

Figure 1. Implied volatility change correlation with the three-month at-the-money option. The figure plots the cross-correlation estimates of percentage implied volatility changes between the reference three-month at-the-money option contract and contracts at other maturities and moneyness. Each line shows the correlation variation across moneyness at one maturity. The solid line corresponds to the correlation estimates with contracts at the same three-month maturity. The two dashed lines correspond to estimates with contracts at the two adjacent maturities at two and six months. The two dash-dotted lines correspond to estimates with contracts at further apart maturities at one and 12 months. (Color figure can be viewed at wileyonlinelibrary.com)

are over 98% for contracts within half a standard deviation ($|x| \leq 0.5$). These extremely high correlation estimates suggest that it is reasonable to apply the local commonality assumption within one standard deviation of moneyness at the same maturity.

The correlation estimates decline as the options become further out-of-the-money. The correlation estimates for contracts at two standard deviations away stay around 75%, suggesting that using one common set of variance/covariance rates will not be sufficient to capture the implied volatility smile across the whole moneyness range.

In Figure 1, the two dashed lines represent the correlation estimates with contracts at the two adjacent two- and six-month maturities, and the two dash-dotted lines represent the correlation estimates with contracts further apart at one-month and 12-month maturities. As the maturity distance increases, the correlation estimates decline further. Nevertheless, if we focus on the at-the-money term structure, the correlation estimates with the two adjacent maturities are at 99% (with two-month maturity) and 97% (with six-month maturity), respectively. The high correlation estimates with the two adjacent maturities suggest that we can apply the commonality assumption to the local at-the-money term structure defined by two adjacent maturities.

In contrast, the correlation estimates with the further-apart maturities drop to 94% with the one-month option and 88% with the 12-month option,

Table I
Correlation Dependence on Maturity and Moneyness Distance

Entries report results from a bivariate regression of the pairwise correlation estimates between daily percentage implied volatility changes of different contracts against the absolute distances between the pair of the contracts in terms of moneyness x and log maturity $\ln \tau$,

$$\rho_{ij} = \alpha + \beta_x |x_i - x_j| + \beta_\tau |\ln \tau_i - \ln \tau_j| + e$$

The regression is performed on the lower triangular of the correlation matrix excluding the diagonal elements.

	α	β_x	β_τ	R^2
Estimates	1.022	−0.078	−0.102	0.776
Std Error	0.005	0.003	0.002	–

highlighting the insufficiency of a one-factor model in capturing global term structure variation over the entire maturity span.

Figure 1 visualizes one slice of the correlation matrix using the three-month at-the-money contract as the reference point. When we consider the entire correlation matrix excluding the diagonal elements, the average correlation estimate is 79%, the median is 81%, the maximum is 99.15%, the minimum is 28.09%, and the standard deviation is 13.55%. Hence, the pairwise correlation estimates can vary over a wide range depending on the particular pair. To formally test our hypothesis that the correlation estimates decline with the absolute distances between the contracts in terms of maturity and moneyness, we perform the following bivariate regression on the lower triangular elements of the correlation matrix excluding the diagonal elements,

$$\rho_{ij} = \alpha + \beta_x |x_i - x_j| + \beta_\tau |\ln \tau_i - \ln \tau_j| + e. \tag{17}$$

Table I reports the regression results. The regression generates a high R^2 of 77.6%. The intercept is close to the null value of 1 when the contract distance is zero. The slope coefficient estimates on both distance measures are strongly negative and highly significant: A one-unit increase in moneyness distance leads to a decrease in correlation of 7.8 percentage points, and a one-unit increase in log maturity distance leads to a decrease in correlation of 10.2 percentage points.

Under the new theory, analyzing the implied volatility surface does not need to start with the full specification of the underlying security price and volatility dynamics, but rather can start with analyzing the co-movement structure on the percentage implied volatility changes of contracts across maturity and moneyness levels. Based on such analysis, one can directly impose cross-sectional factor structures on their variance and covariance, from which the theory generates pricing implications on the level and shape of the implied volatility surface. To gain insights into the factor structure that governs the implied volatility movements across the surface, we perform principal component analysis on the interpolated implied volatility change series.

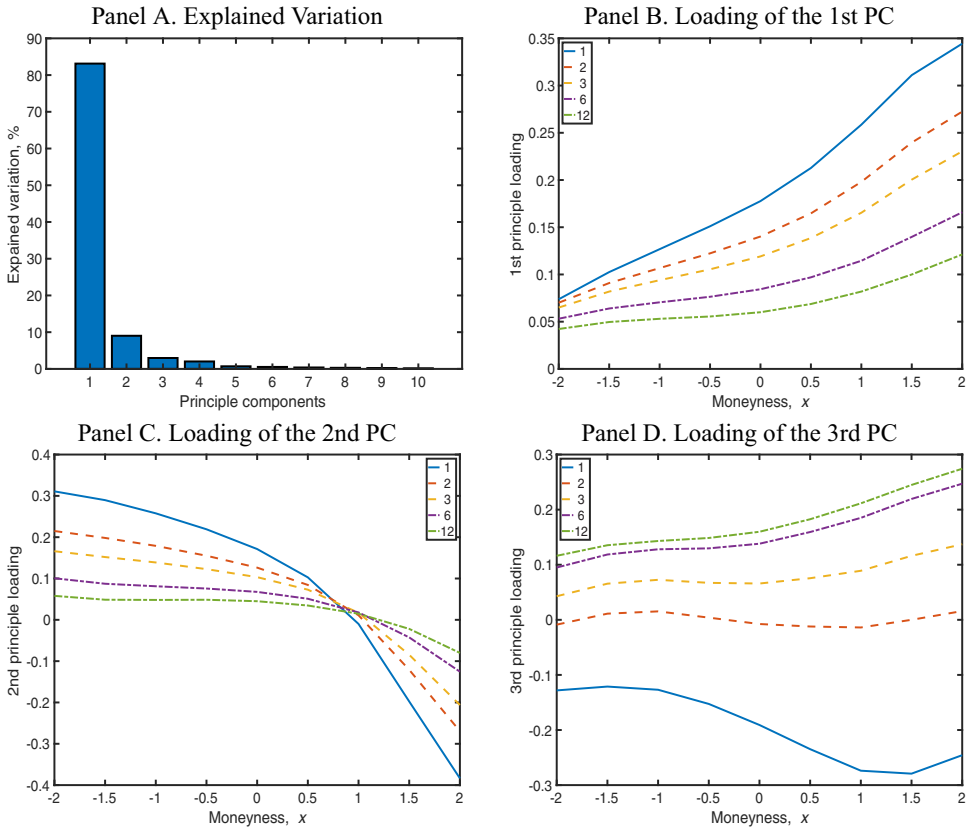

Figure 2. Principal component analysis on implied volatility movements. Panel A uses bar charts to show the explained variation of the top 10 principal components on the 45 interpolated implied volatility change series. Panels B to D plot the loadings of the first, second, and third principal component, respectively, across different moneyness at different maturities, with the solid lines corresponding to the one-month maturity, dashed lines corresponding to the two- and three-month maturities, and dash-dotted lines corresponding to the six- and 12-month maturities. (Color figure can be viewed at wileyonlinelibrary.com)

We estimate the covariance matrix of the 45 series and perform eigenvalue and eigenvector decomposition of the covariance matrix. The eigenvalue vector, up to a normalization, represents the explained variation by each principal component, whereas the eigenvector corresponding to each eigenvalue can be interpreted as the loading of that principal component on the implied volatility surface. Figure 2, Panel A, uses a bar chart to show the percentage variation explained by each of the top 10 principal components. The first component explains 83.13% of the variation, suggesting that the implied volatility surface shares a large proportion of common movements. Figure 2, Panel B, plots the loading of the first principal component, which is universally positive across

all maturity and moneyness levels. The loading estimates are larger at shorter maturities and higher strikes, potentially highlighting their higher variation.

The second principal component explains 9.02% of the variation, which is still highly significant albeit much smaller than the dominant first principal component. Panel C of Figure 2 shows that the loading of this second principal component has a distinct moneyness pattern as the loadings are positive at low-strike regions and negative at high-strike regions, effectively capturing the movement of the implied volatility skew at each maturity.

The third principal component explains 2.97% of the variation. Its effect, as shown in Panel D of Figure 2, is to capture the variation along the term structure. At each moneyness, the loading is negative at short maturities but positive at long maturities.

Taken together, within the confines of our data setting, the first three principal components explain over 95% of the variation on the implied volatility surface. The first principal component captures the movement of the overall implied volatility level. The next two principal components capture the slope changes of the surface along the moneyness and maturity dimensions.

The relative significance of the variations along the two slopes depends on data sampling. The variation along the moneyness dimension tends to be more significant for exchange-traded options, which cover a wide range of moneyness but a narrow range of maturity. For over-the-counter option quotes, which tend to span a narrower range of moneyness but a much wider range of maturity, the term structure variation can become more significant. In general, it is important to realize that the principal component analysis is highly data dependent. To explain the same proportion of variation, one usually needs fewer principal components if the data span a narrower strike or maturity range, but more when the data range is extended. This feature reveals the limitation of efforts to identify global factor structures. A seemingly sufficient factor structure under a certain data setting can often be easily rejected when the data range is expanded.

C. The At-the-Money Implied Variance Term Structure and Its Variation

Table II, Panel A, reports summary statistics for the interpolated floating at-the-money implied volatility levels at the five maturities, $A_t(\tau) \equiv I_t(\tau, 0)$ at $\tau =$ 1, 2, 3, 6, and 12 months. The sample average implied volatility level increases as the time to maturity increases from one month to one year. The one-month at-the-money implied volatility averages 19.6%, very close to the full-sample return standard deviation estimate of 19.5%. The sample average increases with maturity, reaching 21.5% at the one-year maturity.

The standard deviation of the at-the-money implied volatility series declines with maturity from 7.4% at the one-month maturity to 5.6% at the one-year maturity. The range between the historical minimum and maximum also narrows with increasing maturity, consistent with the decreasing standard deviation. The last row of the panel reports the autocorrelation of the floating series. All

Table II
Summary Statistics for At-the-Money Implied Volatility Levels and Daily Changes

The table reports summary statistics for the at-the-money implied volatility level (Panel A), its daily percentage change (Panel B), and the daily percentage changes of at-the-money contracts (Panel C) at maturities of 1, 2, 3, 6, and 12 months. The statistics include the sample average ("Mean"), sample standard deviation ("Stdev"), minimum, maximum, and daily autocorrelation ("Auto") estimates. In Panels B and C, we annualize the mean, standard deviation, minimum, and maximum, and we report the correlation with the index return in the last row ("Corr").

Panel A: At-the-Money Implied Volatility Level, $A_t(\tau)$					
Maturity (τ)	1	2	3	6	12
Mean	0.196	0.199	0.202	0.209	0.215
SD	0.074	0.070	0.068	0.061	0.056
Minimum	0.086	0.097	0.103	0.113	0.120
Maximum	0.736	0.702	0.670	0.587	0.526
Auto	0.984	0.987	0.989	0.993	0.995
Panel B: Daily Log Change in At-the-Money Implied Volatility, $\ln A_{t+1}(\tau)/A_t(\tau)$					
Mean	0.036	0.033	0.033	0.032	0.030
SD	0.912	0.734	0.634	0.470	0.365
Minimum	−82.543	−67.682	−65.370	−55.132	−58.003
Maximum	121.716	99.542	86.031	74.590	55.347
Auto	−0.106	−0.081	−0.063	−0.038	−0.032
Corr	−0.749	−0.776	−0.787	−0.795	−0.781
Panel C: Daily Log Implied Volatility Change for At-the-Money Contracts, $R^A_{t+1}(\tau)$					
Mean	0.133	0.096	0.101	0.102	0.084
SD	0.529	0.407	0.348	0.256	0.200
Minimum	−40.088	−37.851	−37.726	−29.608	−22.723
Maximum	78.000	61.714	52.921	37.376	35.131
Auto	−0.090	−0.052	−0.030	0.015	0.056
Corr	−0.437	−0.474	−0.488	−0.498	−0.473

five series show mean-reversion behavior, stronger at short maturities than at long maturities.

Panel B of Table II reports summary statistics for the daily log change of the at-the-money implied volatility series, $\ln A_{t+1}(\tau)/A_t(\tau)$, at the five maturities. Statistics on the daily changes are annualized. As the floating series show no obvious trend during our sample period, the sample average of the daily change is very small, at about 3% per year across all maturities. The annualized standard deviation estimates of the daily percentage change are very large, from 91.2% at the one-month maturity to 36.5% at the one-year maturity. The minimum and maximum daily percentage changes also show a wide range. The autocorrelation estimates on the daily changes are all negative due to the mean-reverting behavior of the floating series. The last row reports the contemporaneous correlation between the at-the-money volatility change and

the security return. The estimates are strongly negative and of similar absolute magnitudes across all maturities.

Panel C of Table II reports summary statistics for the daily log implied volatility change of the at-the-money contracts, $R^A_{t+1}(\tau) \equiv R_{t+1}(\tau, 0)$, which differ from the daily percentage changes of the floating series in Panel B. Compared to Panel B, the daily log changes on the fixed contracts show a higher annualized sample mean, from 13.3% at the one-month maturity to 8.4% at the one-year maturity. According to the pricing equation in (12) and ignoring the variance risk premium, the positive mean estimates imply an upward-sloping at-the-money implied variance term structure.

The annualized standard deviation estimates in Panel C are smaller than those on the changes of the floating series in Panel B. The estimates range from 52.9% at the one-month maturity to 20% at the one-year maturity. The minimum and maximum also form a narrower range than in Panel B.

The autocorrelation estimates on the implied volatility changes of at-the-money contracts are negative at short maturities, but become positive at long maturities. Thus, mean-reversion in the floating series does not always translate into mean-reversion in the implied volatility changes of the fixed expiry contracts.

The last row reports the contemporaneous correlation with the index return. The estimates are much smaller in absolute magnitude than the correlation estimates on the changes in the floating series. Overall, sliding along the term structure and moneyness leads to large differences in the behavior of the floating and the fixed-contract implied volatility changes.

D. Extract Rate of Change from the At-the-Money Term Structure

Assuming that the proportional movements in implied volatilities are the same for a pair of nearby at-the-money contracts, Proposition 1 allows us to extract the conditional risk-neutral drift of the implied volatility percentage changes (μ_t) within this maturity range from the at-the-money implied variance term structure slope defined by the two nearby contracts. The correlation estimates in Figure 1 show that the implied volatility movements of at-the-money contracts at adjacent maturities are extremely highly correlated, lending support to the local commonality assumption for adjacent maturities. We therefore use the local at-the-money term structure slope defined by two adjacent maturities (τ_i, τ_{i+1}) to estimate the risk-neutral drift at the midpoint of the two maturities,

$$\mu_t(\overline{\tau}_i) = \frac{A_t^2(\tau_{i+1}) - A_t^2(\tau_i)}{2(A_t^2(\tau_{i+1})\tau_{i+1} - A_t^2(\tau_i)\tau_i)}, \qquad \overline{\tau}_i = (\tau_i + \tau_{i+1})/2. \tag{18}$$

Specifically, we use the term structure slope defined by one-month and two-month at-the-money implied volatilities to estimate the drift at 1.5 months, the term structure defined by two-month and three-month at-the-money implied volatilities to estimate the drift at 2.5 months, and so on. We then

Table III
Extract Rate of Change from At-the-Money Implied Variance Term Structure

The table reports summary statistics for the risk-neutral drift (μ_t) extracted from the at-the-money implied variance local term structure defined by the two nearest maturities. The statistics include the sample average ("Mean"), standard deviation ("*SD*"), minimum, maximum, and daily autocorrelation ("Auto") estimates.

Maturity	1	2	3	6	12
Mean	0.243	0.219	0.178	0.104	0.048
SD	0.497	0.367	0.241	0.147	0.081
Minimum	−4.428	−2.756	−0.950	−1.285	−1.047
Maximum	1.186	0.953	0.687	0.390	0.232
Auto	0.932	0.941	0.958	0.969	0.967

linearly interpolate the drift estimates to obtain the estimate at each interpolated maturity.

Table III reports summary statistics for the risk-neutral rate of change (μ_t) estimates. The first row reports the sample average of the risk-neutral estimates, which are larger than the statistical average (Panel C of Table II) at short maturities but smaller at the 12-month maturity. If we regard the average difference as an average risk premium estimate, this result would suggest that on average it is profitable to take short positions in the short-term option contract and long positions in the long-term option contract. Egloff, Leippold, and Wu (2010) obtain a similar conclusion from their estimation of dynamic term structure of variance models on variance swap rates.

The risk-neutral drift estimates show large time series variation. The standard deviation estimates are almost twice as large as the sample mean, and minimum and maximum estimates show that the estimated rate of change can swing from large negative values to large positive values. The autocorrelation estimates in the last row show that the drift estimates are persistent, with autocorrelations from 0.932 at one month to 0.967 at 12 months. If we assume AR(1) dynamics, the autocorrelation estimates imply a half-life of 11 to 23 business days.

To examine the information content of the risk-neutral rate of change estimates in predicting future implied volatility changes, we perform the forecasting regression at each maturity τ,

$$R^A_{t+1}(\tau) = \alpha + \beta \mu_t(\tau) + e_{t+1}, \tag{19}$$

where the log implied volatility change $R^A_{t+1}(\tau)$ is annualized, and $\mu_t(\tau)$ denotes the risk-neutral expected rate of change estimated from the local at-the-money implied volatility term structure at date t. Under the expectation hypothesis that the risk-neutral rate of change μ_t is an unbiased estimator of the future rate of change, we have the null value of $\alpha = 0$ and $\beta = 1$. If we instead expect no information content for the risk-neutral rate of change, the null value for

Table IV

Predict Implied Volatility Changes With the Term Structure Slope

The table reports the coefficient estimates, the absolute Newey-West's t-values (in parentheses), and the R^2s of the following forecasting regression at each maturity τ,

$$R^A_{t+1}(\tau) = \alpha + \beta\mu_t(\tau) + e_{t+1},$$

where the daily log implied volatility change of the at-the-money contract $R^A_{t+1}(\tau)$ is annualized and the risk-neutral expected rate of change $\mu_t(\tau)$ is extracted from the local at-the-money implied variance term structure defined by the two adjacent maturities. The t-values are computed with a lag of one month. For the β estimates, the table reports the t-values for both the null hypothesis of $\beta = 0$ and the null hypothesis of $\beta = 1$.

Maturity	$\widehat{\alpha}$	$H:\alpha=0$	$\widehat{\beta}$	$H:\beta=0$	$H:\beta=1$	R^2, %
1	−0.095	(0.59)	0.935	(2.56)	(0.18)	0.31
2	−0.128	(1.01)	1.027	(3.06)	(0.08)	0.34
3	−0.136	(1.15)	1.332	(3.26)	(0.81)	0.34
6	−0.014	(0.16)	1.115	(2.20)	(0.23)	0.16
12	0.072	(0.99)	0.243	(0.28)	(0.88)	0.00

the slope will be $\beta = 0$. Table IV reports the coefficient estimates, their absolute Newey and West's (1987) t-values computed with 21 lags, and the regression R^2.

The intercept estimates are negative for maturities from one to six months, but positive at the 12-month maturity. Nevertheless, the t-values are small for the intercept estimates and none of them is significantly different from zero. The slope estimates (β) are close to the null value of 1 for all but the 12-month maturity. The t-values against the noninformative null hypothesis of $\beta = 0$ are strongly rejected at one- to six-month maturities, but the t-values against the unbiased null hypothesis of $\beta = 1$ are all very small and cannot be rejected.

The R^2 estimates are all very low, suggesting that predicting the directional movements of the implied volatility for a fixed-expiry contract is inherently difficult. When an investor cannot effectively predict the direction of implied volatility movements, it is prudent to perform a vega hedge and explore opportunities on the shape of the implied volatility smile.

E. The Implied Volatility Smile and its Variation

Table V, Panel A, reports the sample average of the implied volatilities across five selected moneyness levels at each of the five interpolated maturities. At each maturity, the average implied volatility level is higher at lower strikes than at higher strikes, forming the well-known implied volatility skew pattern. Although the skews at most maturities are monotonically downward sloping, the slope tends to be more negative at lower strikes. At very short maturities, the plot shows more curvature and becomes more of a smile pattern as the implied volatility level at very high strikes also becomes higher.

Table V
Mean Implied Volatility Smile and Time Series Variance/Covariance Estimates

The table reports the sample average of the implied volatility levels (Panel A), the covariance (Panel B), and the variance (Panel C) at each selected moneyness (x) and maturity (τ) grid. The variance and covariance are historical time series estimators constructed with a 21-business-day rolling window.

Panel A: Mean Implied Volatility Smile					
$\tau \backslash x$	−2.0	1.0	0	1.0	2.0
1	0.349	0.254	0.196	0.164	0.157
2	0.352	0.260	0.199	0.165	0.153
3	0.354	0.264	0.202	0.166	0.152
6	0.359	0.273	0.209	0.170	0.151
12	0.361	0.278	0.215	0.173	0.152
Panel B: Historical Covariance Estimates γ_t					
1	−0.000	−0.025	−0.045	−0.082	−0.152
2	−0.009	−0.026	−0.038	−0.063	−0.115
3	−0.012	−0.025	−0.033	−0.053	−0.095
6	−0.013	−0.022	−0.025	−0.038	−0.066
12	−0.012	−0.018	−0.019	−0.028	−0.048
Panel C: Historical Variance Estimates ω_t					
1	0.192	0.194	0.280	0.536	1.063
2	0.103	0.118	0.166	0.303	0.630
3	0.074	0.088	0.121	0.214	0.450
6	0.045	0.051	0.066	0.112	0.248
12	0.033	0.033	0.040	0.069	0.149

The new pricing theory links the implied variance smile to the conditional risk-neutral variance of the percentage implied volatility change and its covariance with the security return. Based on the interpolated log implied volatility changes at each maturity moneyness grid $R_t(\tau, x)$ and the index return, we construct the time series estimates of the variance rate (ω_t) and the covariance rate (γ_t) with a 21-business-day rolling window. Panel B of Table V reports the sample average of the historical covariance estimates at each maturity and moneyness grid. The average covariance estimates are negative across all maturities and moneyness, broadly in line with the negatively skewed implied volatility smile. Across maturities, especially around at-the-money, the covariance estimates decline with maturity. Panel C reports the sample average of the variance estimates. The average variance estimates also decline with maturity.

At each maturity, the average covariance estimates are more negative at high strikes than at low strikes, even though the skew tends to be more negative at low strikes. The variance estimates are also higher at high strikes than at low strikes. Although there is systematic institutional demand for out-of-the-money index put options, the demand for out-of-the-money call options tends to

be more retail-driven and less systematic, leading to more idiosyncratic movements for out-of-the-money calls. These larger idiosyncratic movements may help explain the larger absolute magnitude of the variance and covariance estimates at higher strikes. On the other hand, the larger systematic demand for out-of-the-money index put options may have pushed the implied volatility skew steeper at lower strikes, despite the smaller absolute magnitudes of the variance and covariance estimates. This disparity between demand and historical magnitude of variation may present profitable opportunities for investors who can manage the risk appropriately.

F. Extract Variance and Covariance Rates from the Implied Volatility Smile

Figure 1 shows that within one standard deviation of at-the-money ($|x| \leq 1$), log implied volatility changes between the at-the-money contract and other contracts are highly correlated, with correlation estimates over 90%. Based on this evidence, we apply the local commonality assumption within one standard deviation of the at-the-money option ($|x| \leq 1$) and we extract the common risk-neutral variance and covariance rates by performing the cross-sectional regression

$$I_t^2 - A_t^2 = \gamma_t(2z_+) + \omega_t^2(z_+z_-) + e_t. \tag{20}$$

At each date and maturity, we regress the implied variance difference from the at-the-money level against the moneyness measures $2z_+$ and z_+z_-. We constrain the regression to the five moneyness levels within one standard deviation of at-the-money. We can readily convert the moneyness measure x to z_+ and z_- given the implied volatility estimate at each point. We impose zero intercept and constrain the regression coefficient ω_t^2 to be positive. The regression captures the implied variance smile around at-the-money. A negative implied volatility skew will generate a negative covariance estimate γ_t, whereas a positive curvature for the smile leads to a positive variance estimate ω_t^2.

Table VI reports summary statistics for the slope coefficients and the R^2 estimates. Panel A shows that the covariance estimates are universally negative across all maturities and calendar days, suggesting that the implied volatility smiles on the S&P 500 index are persistently negatively skewed. The estimates show high time series persistence. The daily autocorrelation estimates are high, from 0.978 at the one-month maturity to 0.995 at the 12-month maturity.

By choosing options around at-the-money in running the regression, we can regard the regression estimates as reflecting the moment conditions of the at-the-money contract. Comparing the average cross-sectional (CS) regression covariance estimates with the average historical time series (TS) estimates for the at-the-money option in Table V, we find that on average the cross-sectional estimates are more negative than the time series estimates, suggesting that the implied volatility skew tends to be steeper than warranted by the historical behavior of the covariation between the implied volatility and the index return.

Table VI
Cross-Sectional Regression Estimates of Variance and Covariance Rates

The table reports summary statistics for the covariance γ_t, the variance ω_t^2, and the R^2s from cross-sectional regressions of the implied variance spread over the at-the-money level against the moneyness measures $2z_+$ and z_+z_-. The regression is performed for each date and each maturity using the estimates at the five interpolated moneyness levels within $|x| \leq 1$. The statistics include the sample average ("Mean"), sample standard deviation ("*SD*"), minimum, maximum, and daily autocorrelation ("Auto"). For the covariance and variate estimates in Panels A and B, we also report the correlation ("Corr") with the corresponding time series covariance and variance estimates using a 21-business-day rolling window.

Maturity	1	2	3	6	12
Panel A: Covariance Estimates γ_t					
Mean	−0.142	−0.107	−0.090	−0.068	−0.050
SD	0.066	0.046	0.037	0.026	0.018
Minimum	−0.499	−0.360	−0.292	−0.202	−0.132
Maximum	−0.040	−0.038	−0.034	−0.026	−0.018
Auto	0.978	0.986	0.989	0.993	0.995
Corr	0.646	0.660	0.640	0.580	0.525
Panel B: Variance Estimates ω_t^2					
Mean	1.002	0.451	0.270	0.102	0.038
SD	0.461	0.207	0.128	0.056	0.027
Minimum	0.000	0.000	0.000	0.000	0.000
Maximum	2.552	1.065	0.647	0.275	0.240
Auto	0.905	0.933	0.941	0.956	0.958
Corr	−0.141	−0.172	−0.171	−0.130	−0.122
Panel C: Regression R^2s					
Mean	0.999	0.999	0.999	1.000	1.000
SD	0.001	0.001	0.001	0.001	0.001
Minimum	0.983	0.986	0.988	0.984	0.990
Maximum	1.000	1.000	1.000	1.000	1.000
Auto	0.695	0.756	0.738	0.791	0.821

The difference can be interpreted as a risk premium (Driessen, Maenhout, and Vilkov (2009)).

The last row of Table VI, Panel A, reports the cross-correlation between these two sets of estimates. The two series show high degrees of co-movement, with correlation estimates between 0.525 and 0.66. The cross-correlation estimates tend to be higher at short maturities than at longer maturities.

Panel B of Table VI reports summary statistics for the variance estimates identified from the curvature of the implied volatility smile. The average cross-sectional estimates are larger than the average time series variance estimates in Table V at short maturities, but the average magnitudes are similar at the longest (12-month) maturity, suggesting that the implied volatility smile has more curvature at short maturities than at long maturities, and more so than supported by the time series variation of the implied volatility.

The standard deviation estimates show that the curvature estimates vary a lot more than the slope estimates. In particular, the minimum estimates at all maturities are truncated at zero, suggesting that across all maturities there are dates when the interpolated implied variance smiles show little positive curvature. Furthermore, the autocorrelation estimates show that the curvature estimates are less persistent than the slope estimates, and the cross-correlation estimates with the time series variance estimates are negative across all maturities. Taken together, the large standard deviation, the low persistence, and the negative correlation with historical estimates suggest that the cross-sectional identification of the variance via the curvature of the implied variance smile is less reliable than the identification of the covariance via the slope of the smile. When inspecting the time series of the estimates in more detail, we find that during volatile times when the time series variance estimates become large, the smile tends to become highly negatively skewed and the curvature of the smile becomes harder to identify, leading to the negative correlation between the time series and cross-sectional variance estimates.

Panel C of Table VI reports summary statistics for the R^2 estimates of the cross-sectional regressions. The estimates are very high. Across all maturities, the average estimates are 99.9% or higher, and standard deviation estimates are merely 1%. The lowest R^2 estimate is at one-month maturity, and it is still as high as 98.3%. The extremely high R^2s suggest that at least within the narrow range of moneyness ($|x| < 1$), the regression in (20) can capture the implied volatility smile very well, lending support to the local commonality assumption within this moneyness range.

G. Trading the Implied Volatility Smile

The variance and covariance extracted from the cross-sectional regression on the implied volatility smile can differ from the rolling window time series estimates. There are several reasons for the difference. First, the cross-sectional estimates are forward looking, as they are extracted from pricing information today about the future, whereas the time series estimates are backward looking, as they are estimated from recent history. Second, both sets of estimates can have estimation errors, but from different sources. Third, expected values for the two sets of estimates can differ due to risk premium when the price and/or implied volatility process contain random jumps and these jump risks are priced by the market. In this section, we explore the implications of these differences.

G.1. Forecasting Realized (Co)variance with Cross-Sectional and Time Series Estimates

To examine whether the cross-sectional estimates can be combined with the time series estimates to enhance the forecast on future realizations of the variance and covariance, we run the forecasting regression

$$RV_{t+1} = \alpha + \beta_1 CS_t + \beta_2 TS_t + e_{t+1}, \tag{21}$$

Table VII

Predict Realized Variance/Covariance with Cross-Sectional and Time Series Estimators

We predict the one-month-ahead realized covariance γ_t (Panel A) and variance ω_t^2 (Panel B) for the at-the-money implied volatility series with their corresponding cross-sectional estimator (CS) extracted from the implied volatility smile and the one-month rolling historical time series estimator (TS),

$$RV_{t+1} = \alpha + \beta_1 CS_t + \beta_2 TS_t + e_t,$$

where we use RV_{t+1} to denote the one-month-ahead realized estimator. The table reports the regression coefficient estimates, the Newey-West's absolute t-values (in parentheses), and R^2 for each regression. Each row corresponds to one maturity (in months).

Maturity	α		β_1		β_2		R^2, %
			Panel A: Covariance γ_t				
1	−0.011	(1.64)	0.120	(1.79)	0.377	(2.97)	20.23
2	−0.011	(1.65)	0.097	(1.20)	0.433	(3.45)	23.11
3	−0.012	(1.70)	0.070	(0.76)	0.463	(3.84)	24.18
6	−0.009	(1.63)	0.056	(0.58)	0.484	(4.21)	25.28
12	−0.004	(1.11)	0.129	(1.61)	0.466	(4.00)	25.85
			Panel B: Variance ω_t				
1	0.231	(7.02)	−0.004	(0.15)	0.19	(3.72)	3.52
2	0.148	(6.84)	−0.040	(1.07)	0.22	(3.86)	5.40
3	0.116	(6.62)	−0.091	(1.83)	0.25	(4.08)	7.63
6	0.063	(6.47)	−0.160	(2.16)	0.29	(3.95)	10.28
12	0.034	(6.58)	−0.198	(2.26)	0.34	(3.30)	13.46

where RV_{t+1} denotes the realized variance and covariance estimator over the next month, and CS_t and TS_t denote the corresponding cross-sectional and time series estimators, respectively. We perform this regression for both the variance and the covariance at each maturity. Table VII summarizes the regression results for the covariance in Panel A and the variance in Panel B. Entries include the coefficient estimates, the Newey and West's (1987) absolute t-values computed with 21 lags (in parentheses), and the regression R^2s.

The forecasting regressions on the covariance in Panel A generate high R^2 estimates, from 20.23% at the one-month to 25.85% at the 12-month maturity. The coefficients on both the cross-sectional and the time series estimators are positive, but the coefficient estimates are larger, with higher t-values, for the time series estimates. Thus, the time series estimates—based on recent history of the implied volatility changes—represent a good starting point for forecasting how the implied volatility interacts with the index return in the near future.

Panel B shows that the forecasting power on the variance is weaker, which suggests that the variance is less persistent or predictable. The R^2 estimates increase monotonically with maturity, from 3.53% at the one-month maturity

to 13.45% at the 12-month maturity. The coefficient estimates on the historical time series estimator are all positive and statistically significant. The contribution from the cross-sectional estimator is smaller and the coefficient estimates become negative in all cases.

G.2. Risk-Return Trade-Off Trading based on Time Series Forecasts

Given a set of forecasts on future realized variance and covariance and assuming common movements within a moneyness range, our theory generates a breakeven value on the implied volatility smile. The breakeven value represents the valuation of the implied volatility smile based on statistical forecasts on the variance and covariance while assuming zero risk premium. When the observed implied volatility smile differs from this valuation, the difference can be viewed as a potential source of risk premium and a potentially profitable trading opportunity. We examine the profitability using an out-of-sample investment exercise.

Starting on January 3, 2000, at each date and for each option maturity, we use a four-year rolling window to estimate the forecasting relation in (21) to generate out-of-sample forecasts on the covariance $\widehat{\gamma}_t$ and the variance $\widehat{\omega}_t$. With these forecasts, we construct the breakeven implied variance spread at the four moneyness levels $x = \pm 0.5, \pm 1.0$,

$$\widehat{I}_t^2 - A_t^2 = 2\widehat{\gamma}_t z_+ + \widehat{\omega}_t^2 z_+ z_-, \tag{22}$$

where we take the observed at-the-money implied volatility A_t as given and generate fair valuations on the spreads. The two moneyness measures z_+ and z_- also depend on the implied volatility level. We construct the measures using the observed implied volatility level.

At each moneyness, we form vega-neutral spread option portfolios with the at-the-money contract and track the delta-hedged P&L over the next month. As out-of-the-money options tend to be more actively traded than in-the-money options, we form put spreads at low strikes and call spreads at high strikes. For each spread, we normalize the weight on the at-the-money contract to one and take short positions in the out-of-the-money contract to make the spread vega neutral.

We set the weight on each spread based on the difference between the observed implied variance spread $I_t^2 - A_t^2$ from the market and the breakeven valuation $\widehat{I}_t^2 - A_t^2$ from equation (22),

$$w_t = \left(I_t^2 - \widehat{I}_t^2\right)/A_t^2, \tag{23}$$

where the at-the-money implied variance is cancelled out in the numerator and we divide the difference by the at-the-money implied variance level to obtain a proportional number. When the market-observed spread is higher than the fair value, we long $100 \times w_t$ dollars of notional of the at-the-money contract and short the contract at the corresponding moneyness to make it

vega neutral. We hold the spread for 21 days while performing daily delta hedge with the underlying index futures. In option investments, because of the small value of the option contract relative to the potentially large risk exposures, investors often measure investments in terms of notional exposures rather than the money spent or received from buying or selling the option. Investing $100 \times w_t$ notional dollars is equivalent to normalizing the index level to \$100 and investing w_t shares in the normalized contract.

Table VIII reports summary statistics for the investment weights. Each panel reports one summary statistic. Within each panel, each column represents one moneyness level x, and each row represents one time to maturity τ, represented in months. The average weights on the put spreads at $x = -0.5$ and -1 are positive and on the call spreads at $x = 0.5$ and 1 are negative. The strategy is, on average, short out-of-the-money puts and long out-of-the-money calls, benefiting from the observation that on average the negative implied volatility skews are steeper than warranted by the historical covariance estimates. The average weights are larger in absolute magnitude for further out-of-the-money spreads at $|x| = 1$ than at $|x| = 0.5$, as are the standard deviation estimates.

Nevertheless, the minimum weights are negative and the maximum weights are positive on all spreads, suggesting that the investments based on the time series forecasts can switch signs despite the average bias. Finally, the daily autocorrelation estimates in Panel E show that the investment weights are highly persistent. This high persistence suggests low turnover for the strategy.

In the three panels in Table IX, we track the delta-hedged P&L of each spread portfolio and report its annualized mean, annualized standard deviation, and annualized information ratio. Table IX is similar to Table VIII in that within each panel, each column represents one moneyness level and each row represents one time to maturity. In addition to the four moneyness levels at each maturity, we add a column labelled "All," which reports the P&L for investing in all four spreads at each maturity.

The sample averages of the P&L on all spread investments are positive. The investments show larger risk (standard deviation) at shorter maturities than at longer maturities, and also at further out-of-money at $|x| = 1$ than at $|x| = 0.5$. The average P&L also varies with maturity and moneyness, but not as much as the risk. The net result is that the information ratios are higher for investments at longer maturities than at shorter maturities. For example, combining the four spreads at the one-month maturity generates an annualized information ratio of 0.52. The information ratio becomes increasingly larger as the option maturity increases, reaching as high as 3.81 for investing in the four spreads at the 12-month maturity. Several factors contribute to the low profitability of the trades on short-term contracts. First, short-term options are more sensitive to price and implied volatility jumps, a component absent from our pricing relation. Second, the forecasting results in Table VII show that the R^2 of the forecasting regression is lower at short maturities. Thus, our forecasts of the breakeven spreads are also less accurate, further reducing trading performance. By contrast, for long-dated contracts, the variance and covariance of the implied volatility series dominate the P&L variation of the

Table VIII

Summary Statistics on the Risk-Return Strategy Investment Weights

The table reports summary statistics for the investment weights on each option spread for the risk-return strategy, where the weights are determined by the difference between the observed spread and the breakeven spread computed based on the time series forecasts of the future realized variance and covariance of the implied volatility changes.

$\tau \backslash x$	−1.0	−0.5	0.5	1.0
		Panel A: Sample Average		
1	0.481	0.178	−0.117	−0.189
2	0.455	0.175	−0.117	−0.194
3	0.440	0.174	−0.120	−0.201
6	0.416	0.175	−0.131	−0.221
12	0.353	0.158	−0.132	−0.227
		Panel B: Standard Deviation		
1	0.149	0.056	0.038	0.068
2	0.140	0.053	0.037	0.066
3	0.142	0.054	0.038	0.066
6	0.148	0.057	0.038	0.064
12	0.156	0.058	0.038	0.061
		Panel C: Minimum		
1	−0.132	−0.071	−0.278	−0.472
2	−0.355	−0.168	−0.223	−0.368
3	−0.592	−0.256	−0.222	−0.373
6	−0.762	−0.305	−0.221	−0.375
12	−0.802	−0.208	−0.232	−0.391
		Panel D: Maximum		
1	0.986	0.386	0.037	0.084
2	0.945	0.343	0.103	0.189
3	0.921	0.337	0.163	0.272
6	0.847	0.319	0.165	0.245
12	1.209	0.319	0.076	0.116
		Panel E: Daily Autocorrelation		
1	0.946	0.933	0.946	0.957
2	0.962	0.946	0.958	0.969
3	0.969	0.956	0.964	0.972
6	0.979	0.975	0.976	0.980
12	0.977	0.983	0.983	0.984

option investments, and our forecasting regression generates a decent forecast of what the future variance and covariance will be. As a result, investing based on the difference between the observed smile and the breakeven valuation—constructed with statistical moment condition forecasts—can generate high information ratios.

Table IX
Out-of-Sample Option Investment P&L Statistics from the Risk-Return Trade-Off Strategy

The table reports the annualized mean (Panel A), the annualized standard deviation (Panel B), and the annualized information ratio (Panel C) of the out-of-sample option investment P&L from the risk-return trade-off strategy on each option spread. Within each panel, each row corresponds to one maturity, and each column to one moneyness level. The "All" column reports the P&L for investing in all four moneyness levels at each maturity. Each investment is made vega neutral with the at-the-money contract of the same maturity. The investment in each contract at each point in time is based on the difference between the observed implied variance and its forecasted breakeven level.

$\tau \backslash x$	−1.0	−0.5	0.5	1.0	All
			Panel A: Annualized Mean		
1	0.582	0.060	0.060	0.182	0.884
2	1.109	0.195	0.104	0.305	1.713
3	1.053	0.191	0.105	0.311	1.660
6	0.822	0.160	0.101	0.304	1.387
12	0.517	0.109	0.082	0.254	0.962
			Panel B: Annualized Standard Deviation		
1	1.298	0.258	0.166	0.524	1.689
2	0.624	0.107	0.060	0.183	0.888
3	0.417	0.077	0.047	0.147	0.638
6	0.238	0.047	0.033	0.109	0.401
12	0.146	0.029	0.023	0.079	0.252
			Panel C: Annualized Information Ratio		
1	0.45	0.23	0.36	0.35	0.52
2	1.78	1.82	1.75	1.67	1.93
3	2.52	2.47	2.25	2.11	2.60
6	3.46	3.39	3.06	2.79	3.46
12	3.53	3.76	3.59	3.21	3.81

The difference between the market observation and the breakeven valuation reflects a potential source of risk premium. By taking positions proportional to the difference as in (23), the strategy strives to earn the risk premium while hedging away the delta and vega risk. To highlight the risk premium positioning nature of the strategy, we label the strategy as the *risk-return trade-off* strategy, and contrast it in the next subsection with the statistical arbitrage strategies commonly proposed based on cross-sectional fitting of no-arbitrage models.

Many papers have documented risk premiums on option investments, such as the early documentation of option returns by Coval and Shumway (2001), the average negative variance risk premium documented by Carr and Wu (2009), and the particularly expensive nature of out-of-the-money stock index puts documented by Wu (2006) and Bondarenko (2014). By linking the shape of the local smile to the variance and covariance of the implied volatility of the

underlying option contract, our new theory allows us to identify a source of risk premium based on the difference between the cross-sectional and time series estimates of these variance and covariance rates.

G.3. Statistical Arbitrage Trading Based on Cross-Sectional Fitting

A popular investment strategy commonly associated with no-arbitrage models is statistical arbitrage trading based on the pricing errors of a cross-sectionally fitted no-arbitrage model.[5] The idea underlying this strategy is to regard the pricing errors as temporal market mispricing and to form portfolios that are neutral to the risk factors of the model, so that the strategy does not take on the risk premiums of these risk factors, but bets purely on the mean-reversion of the pricing errors.

To highlight the difference between the two types of investment strategies, we consider a statistical arbitrage trading strategy within the same context of our risk-return trade-off investment exercise. Specifically, at each date and maturity, we perform the following cross-sectional regression on the implied volatility smile within $|x| \leq 1$:

$$S_t = \gamma_t(2z_+) + \omega_t^2(z_+z_-) + e_t. \tag{24}$$

We have used this regression in equation (20) to estimate the risk-neutral variance and covariance by assuming local commonality for moneyness within $|x| \leq 1$. If our assumption is correct, the pricing errors from this cross-sectional regression (e_t) can be regarded as temporary market mispricing and can be treated as statistical arbitrage opportunities. We set the weight invested in each spread proportional to the pricing error,

$$w_t = 10e_t/A_t^2. \tag{25}$$

Different from the risk-return trade-off strategy, the weight for this statistical arbitrage strategy does not depend on time series forecasts, but rather relies purely on the pricing errors from the cross-sectional fitting. Although the two strategies invest in the same spreads, their investment weights are based on completely different sources of information and assumptions.

As shown in Table VI, the local commonality assumption works well within the moneyness range, and this regression fits the smile extremely well. As a result, the pricing errors are very small. We scale up the error by a magnitude of 10 in equation (25), so that the absolute magnitudes of the weights are in a similar range as those from the risk-return trade-off strategy.

Table X reports summary statistics for the investment weights on the statistical arbitrage strategy. Via scaling, we have made the standard deviation and the minimum-maximum range of the investment weights to be of similar

[5] Duarte, Longstaff, and Yu (2007) describe the performance of several common statistical arbitrage strategies in the fixed income market. Bali, Heidari, and Wu (2009) elaborate on the construction and performance of a statistical arbitrage trading strategy on the swap rate curve based on pricing errors from multifactor dynamic term structure models.

Table X
Summary Statistics for the Statistical Arbitrage Strategy Investment Weights

The table reports summary statistics for the investment weights on each option spread for the statistical arbitrage strategy, where the weights are determined by the cross-sectional fitting errors on the implied variance smile at each maturity.

$\tau \backslash x$	−1.0	−0.5	0.5	1.0
		Panel A: Sample Average		
1	−0.004	−0.012	−0.028	0.002
2	−0.008	0.005	−0.021	0.003
3	−0.013	0.019	−0.026	0.007
6	−0.016	0.027	−0.047	0.019
12	−0.015	0.011	−0.056	0.016
		Panel B: Standard Deviation		
1	0.053	0.135	0.081	0.074
2	0.049	0.137	0.086	0.074
3	0.047	0.137	0.087	0.077
6	0.038	0.121	0.080	0.070
12	0.038	0.108	0.076	0.063
		Panel C: Minimum		
1	−0.457	−0.883	−0.408	−0.853
2	−0.375	−0.723	−0.350	−0.619
3	−0.316	−0.559	−0.361	−0.603
6	−0.298	−0.447	−0.339	−0.570
12	−0.382	−0.386	−0.511	−0.596
		Panel D: Maximum		
1	0.227	0.648	0.372	0.240
2	0.177	0.574	0.325	0.196
3	0.142	0.562	0.361	0.217
6	0.133	0.443	0.460	0.184
12	0.109	0.382	0.312	0.169
		Panel E: Daily Autocorrelation		
1	0.713	0.746	0.733	0.707
2	0.786	0.817	0.826	0.855
3	0.801	0.858	0.855	0.886
6	0.806	0.906	0.894	0.913
12	0.813	0.933	0.914	0.895

magnitudes to those for the risk-return strategy in Table VIII. Nevertheless, we observe several differences. First, different from the large average bias on the risk-return investment weights, the investments weights for the statistical arbitrage strategy, which are determined by the fitting errors of the cross-sectional pricing relation, do not have any large average biases. The lack of

Table XI

Out-of-Sample Option Investment P&L Statistics from the Statistical Arbitrage Strategy

The table reports the annualized mean (Panel A), the annualized standard deviation (Panel B), and the annualized information ratio (Panel C) of the out-of-sample option investment P&L from the risk-return trade-off strategy on each option spread. Within each panel, each row corresponds to one maturity, and each column to one moneyness level. The "All" column reports the P&L for investing in all four moneyness levels at each maturity. Each investment is made vega neutral with the at-the-money contract of the same maturity. The investment in each contract at each point in time is based on the fitting errors from a cross-sectional regression on the implied variance smile.

$\tau \backslash x$	−1.0	−0.5	0.5	1.0	All
			Panel A: Annualized Mean		
1	0.075	−0.019	0.015	0.008	0.079
2	−0.006	0.037	0.019	−0.014	0.035
3	−0.026	0.050	0.024	−0.015	0.032
6	−0.027	0.032	0.035	−0.019	0.021
12	−0.023	0.010	0.037	−0.010	0.014
			Panel B: Annualized Standard Deviation		
1	0.220	0.249	0.113	0.218	0.369
2	0.138	0.111	0.048	0.085	0.061
3	0.103	0.095	0.041	0.075	0.047
6	0.050	0.059	0.034	0.065	0.036
12	0.029	0.035	0.021	0.040	0.020
			Panel C: Annualized Information Ratio		
1	0.31	−0.08	0.13	0.03	0.23
2	−0.04	0.34	0.38	−0.15	0.53
3	−0.25	0.53	0.58	−0.19	0.66
6	−0.54	0.55	1.04	−0.28	0.61
12	−0.79	0.28	1.77	−0.24	0.68

bias suggests that our pricing relation performs uniformly well within our constrained moneyness range. Second, while the risk-return investment weights tend to be larger in absolute magnitude and in variation for further out-of-the-money spreads, the statistical arbitrage investment weights are much more uniform. Third, different from the high persistence of the risk-return investment weights, the daily autocorrelation estimates for the statistical arbitrage investment weights are much smaller, between 0.73 and 0.91, implying a half-life of merely three to eight days under the AR(1) dynamics assumption. As the purpose of the statistical arbitrage strategy is to benefit from temporal mispricing, the investment weights are unlikely to persist in one direction. Thus, the strategy necessarily involves higher turnover.

Table XI reports summary statistics for the investment P&L of the statistical arbitrage strategy. Compared to corresponding investment performance from the risk-return strategy in Table IX, the statistical arbitrage strategy does not work well at the individual spread level. The average P&L estimates

on the individual spread investments are small and in many cases negative. Nevertheless, combining the four spreads at each maturity leads to positive information ratio estimates across all maturities, albeit lower than those from the risk-return trade-off strategy in Table IX.

The investment performance difference between the two strategies has less to do with one strategy being superior to the other, and more to do with the particular setting of the exercise and the different applications of the two strategies. The key for a profitable risk-return trade-off strategy is to identify a robust source of risk premium and accurately predict its time-variation. The setting of our investment exercise is tilted toward this purpose: The exercise focuses on a narrow range of moneyness at each maturity so we can apply the local commonality assumption and link the implied volatility smile within this range to a common set of variance and covariance rates. By contrast, the statistical arbitrage strategy does not take on systematic risks to benefit from risk premiums, but rather bets on the reversion of the remaining pricing errors. As a result, the investment performance of the statistical arbitrage strategy depends not on the accurate identification of the risk premiums, but on the existence of multiple sources of independent, highly mean-reverting pricing errors. Given our focus on a narrow moneyness range and the selection of a limited number of contracts, it is understandable that the statistical arbitrage strategy does not perform as well. Its performance can be improved by considering a large cross section, for instance, by calibrating a multifactor stochastic volatility model on the entire implied volatility surface and forming a portfolio of many options neutralized to the underlying risk factors.[6]

In line with the different focus of the two strategies, we also observe very different cross-correlation behaviors for the P&L series on different option spreads. For the risk-return investment strategy, because all of the spreads are betting on a similar source of risk premium, the P&L series on the four option spreads at each maturity tend to be positively correlated. The correlation estimates on the P&L series average from 36% at the one-month maturity to over 80% for maturities of three months and longer. As a result of the positive correlation, combining the four spreads at each maturity does not increase the information ratio much. By contrast, because the pricing errors from the cross-sectional fitting are supposed to be idiosyncratic, the P&L series from the statistical arbitrage strategy tend to be low or negatively correlated. Combining the different series into one portfolio tends to increase the information ratio much more significantly.

The out-of-sample exercise is very stylized. The allocation weights are simply set proportional to the pricing deviations without further optimization, and the exercise does not adjust for transaction costs. In practical implementations, trading costs tend to be a smaller concern for the risk-return strategy due to low turnover. By contrast, the statistical arbitrage strategy is more suited

[6] One can estimate a multifactor stochastic volatility model (e.g., Carr and Wu (2017b)) while maintaining time series consistency, or can use a simpler model (e.g., Bakshi, Cao, and Chen (1997)) but with frequent recalibration.

for market makers who receive the bid-ask spread in entering the trades. The purpose of a market maker is not to make money on systematic risk premiums, but rather to receive bid-ask spreads as compensation for providing liquidity while maintaining a well-hedged inventory. A well-specified bottom-up no-arbitrage model can be the basis for managing the risk of such an inventory.

The exercise above highlights the difference, from the perspective of investment applications, between traditional bottom-up option pricing models that emphasize valuation consistency across a large cross section of interrelated derivative securities and our new pricing framework that emphasizes the risk-return trade-off on a particular contract. This distinction, of course, is not absolute. At least in principle, it is possible to identify risk premiums on the underlying risk factors of a bottom-up option pricing model via an estimation approach that includes both the time series and the cross section (e.g., Carr and Wu (2017b)). It is equally possible to derive cross-sectional constraints on our new pricing framework by making global factor or commonality assumptions on the moment conditions. Nevertheless, the two approaches allow one to examine the valuation of an option contract from two different perspectives and to draw potentially different insights.

V. Implications, Limitations, and Risk Representation

Traditional derivative pricing models specify the underlying security price dynamics so that one can take expectations on future payoffs and discount them to obtain the present value. The approach is analogous to classic discounted cash flow valuation of primary securities. By contrast, our new theory builds on the analysis of an option investment's instantaneous return, and is therefore closer to traditional risk-return analysis such as the classic capital asset pricing model. Under our assumptions and with discrete time notation, the time-t conditional joint risk-neutral return distribution of the underlying security and the option's implied volatility can be written as

$$\begin{bmatrix} R^S_{t+1} \equiv \dfrac{\Delta S_{t+1}}{S_t} \\ R^I_{t+1} \equiv \dfrac{\Delta I_{t+1}}{I_t} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ \mu_t \end{bmatrix}, \begin{bmatrix} \sigma_t^2 & \gamma_t \\ \gamma_t & \omega_t^2 \end{bmatrix} \right). \tag{26}$$

The mean-variance structure determines the pricing of the option contract according to

$$I_t^2 = \sigma_t^2 + 2\tau\mu_t I_t^2 + 2\gamma_t z_+ + \omega_t^2 z_+ z_-. \tag{27}$$

Pricing multiple option contracts on the same underlying security amounts to expanding the mean-variance structure in (26) to include a vector of returns on multiple option contracts. Pricing options on a large universe of underlying securities amounts to expanding the structure further to include all of the securities of interest, analogous to traditional mean-variance analysis on primary securities.

Under this new mean-variance derivative pricing framework, the pricing and investment decisions on derivatives are as good as the forecasts for the mean vector and the covariance matrix. Investigating how to obtain robust risk-neutral estimates to generate robust option pricing, and how to obtain accurate statistical forecasts to make profitable investment decisions represent interesting future research opportunities. To make investment decisions on a large universe of financial securities, the industry has developed large factor structures (e.g., BARRA) to enhance the robustness of the covariance matrix estimates. Analogous structures can be developed to construct the covariance matrix of the implied volatility changes for pricing, for risk management, and for investment decisions on derivatives.

Discounted cash flow valuations and risk-return analyses are different sides of the same coin. They do not directly compete, but rather offer insights from complementary perspectives. Our proposed framework for mean-variance analysis on derivatives serves a similar complementary role to traditional derivative pricing models.

A. Implications

In discounted cash flow valuations for stocks, one usually makes an assumption on some permanent growth rate for projections of cash flows infinitely into the future and some convergence of risk to some steady level. These far-into-the future assumptions are impossible to be accurate, but can nevertheless have a sizable impact on the valuations. In derivative pricing, the analogy is usually some stationarity assumption on stochastic volatilities. Stationarity assumptions seem innocuous and reasonable, but they can often generate conflicting implications.

Giglio and Kelly (2018) show that in a wide range of applications, commonly used stationary models calibrated to short maturities generate variation in long-dated contracts that is much smaller than observed. Carr and Wu (2003) observe that as long as stochastic volatility is stationary, the central limit theorem should hold and the implied volatility smile at long maturities should flatten out, but the implied volatility skews on the S&P 500 index options do not become flatter but rather become steeper as maturity increases to as long as five years (Foresi and Wu (2005)). The steep skew at long maturities is consistent with the observed large variation in the implied volatility, but neither the large variation nor the steep skew match the implications of a stationary stochastic volatility model estimated with data at shorter maturities.

By linking the pricing to the current variation in the implied volatility of the same contract, our new theory does not rely on such stationarity assumptions, but depends more on direct observations of its own variation. What may appear puzzling from the perspective of long-run stationarity no longer appears so when considered from its own observed variation.

The currency options market has also experienced a similar transition in perspectives. When broker dealers are asked to offer quotes on long-dated currency option contracts, they often use short-dated exchange-traded options

as a starting point and perform extrapolation. In early years, the common practice was to flat-extrapolate the at-the-money option implied volatility from short to long maturities when there was no additional information about the long-run behavior. To extrapolate the implied volatilities of out-of-the-money options, broker dealers and quantitative researchers recognized the effect of the central limit theorem, which led them to quote increasingly flatter implied volatility smiles as the extrapolated maturity increased. Years of investment experiences, however, has made them realize the inadequacy and inconsistency of such an extrapolation approach. Since the early 2000s, the common practice has shifted to extrapolating the entire implied volatility smile shape instead of just the at-the-money implied volatility level. The new extrapolation approach implies that the risk-neutral return skewness and kurtosis stay the same as the conditional horizon increases. This extrapolation seems to run against the central limit theorem. However, out of the two approaches, only the more recent approach is consistent with the implication of our new theory: If one flat-extrapolates the at-the-money implied volatility from one year to 10 years, both the variation and the covariation of the 10-year implied volatility will be identical to those of the one-year implied volatility, and our new pricing theory suggests that the two maturities must share the same implied volatility smile shape as well. The implication of the central limit theorem based on stationarity assumptions cannot kick in as long as the variation in the implied volatility remains large.

B. Limitations

Mean-variance analysis has been the pillar of modern finance. Our new theory can become the stepping stone for expanding mean-variance analysis to derivative securities. However, the mean-variance framework fails to account for the presence of rare but large events (jumps). The literature includes numerous efforts to integrate jumps into mean-variance analysis for both pricing and investment decisions on primary securities. Future research could examine how to integrate jumps into our new framework for derivative pricing.

In many cases, incorporating jumps results in complications that make the analysis lose its original clarity and appeal. In such cases, instead of complicating the analysis via integration, one can follow the same spirit that our new theory advocates and decentralize, analyzing normal-day behavior via the simple mean-variance structure while analyzing rare, large events separately via scenario analysis and stress tests.

C. Risk Representation

In deriving our pricing equation for a vanilla option contract, we use the BMS model to transform the option price into the BMS implied volatility as a representation of the option contract's main risk sources, analogous to the common practice in the fixed income market of transforming bond prices into yield to maturities to highlight the underlying interest rate risk.

Traditional option pricing models represent the risk of many option contracts via a few global risk factors. To generate localized pricing implications, we require that the risk representation be local to the particular option contract. Just as yield to maturity is local to the particular coupon bond, so is implied volatility local to the particular option contract. Accordingly, in choosing a particular form of risk representation for our purpose, the first requirement is that the transformation be a local, unique, monotone, and one-to-one mapping to the option price. For this purpose, the local volatility model of Dupire (1994) or stochastic volatility models such as Heston (1993) are not good candidates.

The second requirement is that the transformed quantities become more stable and more comparable across contracts and over time than the raw contract prices. For bonds, if the instantaneous interest rate is constant, yield to maturity across all bonds will be the same. For options, if the security price follows a diffusive process with a constant instantaneous return volatility, the BMS implied volatility will be the same across all option contracts underlying the same security. This analogous feature allows the yield and the implied volatility to be both somewhat stable and comparable across contracts and over time. Just as yields of nearby maturities tend to be similar and move together by nature of arbitrage trading, so do implied volatilities of option contracts at nearby strikes and maturities.

A third requirement is ease of interpretation in that the transformed quantities have some specific and intuitive economic meaning, at least under certain model environments. The yield to maturity has the economic interpretation of being the return to the bond investment if one buys the bond at the current price and holds the bond to maturity assuming no default. The BMS implied volatility has the economic interpretation of being the security return volatility under the BMS model environment. More generally, under continuous price dynamics, the BMS implied variance represents the risk-neutral expected value of the integrated future instantaneous return variance weighted by the BMS gamma along the sample path.

Nevertheless, it is important to emphasize that although the BMS formula represents a very intuitive and simple representation of an option contract's underlying risk sources, one can explore other transformations that may represent the underlying risk sources even better. As an example, a recent working paper by Shue and Townsend (2018) suggests that people tend to think additively instead of multiplicatively, so that stock price movements look more arithmetic than geometric. If this is indeed the case, an alternative representation would be the Bachelier's (1900) model, under which one measures volatility on price changes instead of percentage returns.

Finally, this paper uses the option contract as an example and uses the well-adopted BMS implied volatility as the transformation. For other financial contracts, it is natural to find the risk representation that best suits the particular contracts under consideration. As we discuss above, if the analysis is on coupon bonds, the yield-to-maturity transformation will be the natural choice (Carr and Wu (2017a)).

VI. Concluding Remarks

We develop a new option pricing framework based on the P&L attribution analysis of option investments. The analysis starts with the Black-Merton-Scholes option pricing equation and attributes the instantaneous return of an option investment to calendar decay, to changes in the underlying security price, to changes in the option's implied volatility, and to higher order effects. Taking risk-neutral expectation and applying dynamic no-arbitrage constraints results in a pricing relation that links the option's fair implied volatility level to the implied volatility's own expected direction of movement, its variance, and its covariance with the security return. The valuation does not need to specify where the first- and second-moment conditions come from or how they vary in the future, thus allowing us to make top-down, localized valuations based on what we know best about the particular contract.

The top-down perspective does not preclude us from making either local commonality assumptions on the implied volatility co-movements across nearby option contracts, or global commonality assumptions based on principal component analysis or some other factor structures. Imposing these commonality assumptions on our pricing equation leads to cross-sectional pricing implications either for a select number or range of contracts or for the entire implied volatility surface.

By shifting the focus from terminal payoffs to short-term P&L fluctuations, our new theory not only provides simple and flexible pricing solutions for derivative contracts of interest, but also tightly links the pricing to risk management practices, and provides a foundation for expanding classic mean-variance risk-return analysis to derivative contracts.

In future research, one could expand existing return factor structures on primary securities by developing analogous factor structures on the implied volatility changes of the underlying option contracts. The joint return/implied volatility factor structure has direct pricing implications for the option contracts, making it suitable as the foundation for investment analysis on the expanded universe that includes both primary securities and derivative contracts.

Initial submission: April 23, 2018; Accepted: October 16, 2019
Editors: Stefan Nagel, Philip Bond, Amit Seru, and Wei Xiong

Appendix: BMS Sensitivities

The standard greeks of the BMS model are well known and are available in standard textbooks, such as Chapter 17 in Hull (2009). A call option's BMS delta is given by

$$B_S = N(d_1), \quad d_1 = \frac{\ln S_t/K + \frac{1}{2} I_t^2 \tau}{I_t \sqrt{\tau}} = -\frac{z_-}{I_t \sqrt{\tau}}, \tag{A1}$$

where $N(\cdot)$ is the cumulative normal function. The BMS gamma is

$$B_{SS} = \frac{n(d_1)}{S_t I_t \sqrt{\tau}}, \tag{A2}$$

where $n(\cdot)$ denotes the probability density function of the standard normal variable. The cash gamma $B_{SS} S_t^2$ is therefore

$$B_{SS} S_t^2 = \frac{S_t n(d_1)}{I_t \sqrt{\tau}}. \tag{A3}$$

All other sensitivity measures of interest can be represented in terms of the cash gamma. The BMS theta is

$$B_t = -\frac{1}{2} S_t n(d_1) I_t / \sqrt{\tau} = -\frac{1}{2} I_t^2 B_{SS} S_t^2. \tag{A4}$$

The BMS vega is

$$B_I = S_t \sqrt{\tau} n(d_1), \tag{A5}$$

so the cash vega $B_I I_t$ can be represented by

$$B_I I_t = S_t I_t \sqrt{\tau} n(d_1) = I_t^2 \tau B_{SS} S_t^2. \tag{A6}$$

Vanna and volga are not in standard textbooks but can be derived just as easily. The vanna is

$$B_{IS} = \frac{\partial B_I}{\partial S_t} = \sqrt{\tau} n(d_1) - d_1 n(d_1) / I_t, \tag{A7}$$

so the dollar vanna $B_{IS} I_t S_t$ can be represented by

$$\begin{aligned} B_{IS} I_t S_t &= S_t n(d_1)(I_t \sqrt{\tau} - d_1) = \frac{S_t n(d_1)}{I_t \sqrt{\tau}} \left(\ln K / S_t + \frac{1}{2} I_t^2 \tau \right) \\ &= B_{SS} S_t^2 z_+. \end{aligned} \tag{A8}$$

Finally, the volga is

$$B_{II} = \frac{\partial B_I}{\partial I_t} = S_t \sqrt{\tau} n(d_1)(-d_1) \left(-\frac{\ln S_t / K}{I_t^2 \sqrt{\tau}} + \frac{1}{2} \sqrt{\tau} \right), \tag{A9}$$

so the dollar volga $B_{II} I_t^2$ can be represented by

$$B_{II} I_t^2 = S_t n(d_1)(-d_1) \left(-\ln S_t / K + \frac{1}{2} I_t^2 \tau \right) = B_{SS} S_t^2 z_- z_+. \tag{A10}$$

REFERENCES

An, Byeong-Je, Andrew Ang, Turan G. Bali, and Nusret Cakici, 2014, The joint cross section of stocks and options, *Journal of Finance* 69, 2279–2337.

Arrow, Kenneth J., and Gerard Debreu, 1954, Existence of an equilibrium for a competitive economy, *Econometrica* 22, 265–290.

Arslan, Melih, Gilbert Eid, Jean El Khoury, and Jason Roth, 2009, The gamma vanna volga cost framework for constructing implied volatility curves, Working paper, Deutsche Bank.

Avellaneda, Marco, and Yingzi Zhu, 1998, A risk-neutral stochastic volatility model, *International Journal of Theoretical and Applied Finance* 1, 289–310.

Bachelier, Louis, 1900, Theorie de la speculation, Ph.D. thesis Paris.

Bakshi, Gurdip, Charles Cao, and Zhiwu Chen, 1997, Empirical performance of alternative option pricing models, *Journal of Finance* 52, 2003–2049.

Bakshi, Gurdip, and Dilip Madan, 2000, Spanning and derivative-security valuation, *Journal of Financial Economics* 55, 205–238.

Bali, Turan, Massoud Heidari, and Liuren Wu, 2009, Predictability of interest rates and interest-rate portfolios, *Journal of Business and Economic Statistics* 27, 517–527.

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

Black, Fisher, and Myron Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* 81, 637–654.

Bondarenko, Oleg, 2014, Why are put options so expensive? *Quarterly Journal of Finance* 4, 1–50.

Boyer, Brian H., and Keith Vorkink, 2014, Stock options as lotteries, *Journal of Finance* 69, 1485–1527.

Byun, Suk-Joon, and Da-Hea Kim, 2016, Gambling preference and individual equity option returns, *Journal of Financial Economics* 122, 155–174.

Carr, Peter, and Liuren Wu, 2003, Finite moment log stable process and option pricing, *Journal of Finance* 58, 753–777.

Carr, Peter, and Liuren Wu, 2009, Variance risk premiums, *Review of Financial Studies* 22, 1311–1341.

Carr, Peter, and Liuren Wu, 2010, Stock options and credit default swaps: A joint framework for valuation and estimation, *Journal of Financial Econometrics* 8, 409–449.

Carr, Peter, and Liuren Wu, 2017a, Decomposing long bond returns: A decentralized modeling approach, Working paper, New York University and Baruch College.

Carr, Peter, and Liuren Wu, 2017b, Leverage effect, volatility feedback, and self-exciting market disruptions, *Journal of Financial and Quantitative Analysis* 52, 2119–2156.

Castagna, Antonio, and Fabio Mercurio, 2007, The vanna-volga method for implied volatilities, *Risk* , 1, 106–111.

Coval, Joshua D., and Tyler Shumway, 2001, Expected option returns, *Journal of Finance* 56, 983–1009.

Daglish, Toby, John Hull, and Wulin Suo, 2007, Volatility surfaces: Theory, rules of thumb, and empirical evidence, *Quantitative Finance* 7, 507–524.

Driessen, Joost, Pascal J. Maenhout, and Grigory Vilkov, 2009, The price of correlation risk: Evidence from equity options, *Journal of Finance* 64, 1377–1406.

Duarte, Jefferson, Francis A. Longstaff, and Fan Yu, 2007, Risk and return in fixed income arbitrage: Nickels in front of a streamroller, *Review of Financial Studies* 20, 769–811.

Dupire, Bruno, 1994, Pricing with a smile, *Risk* 7, 18–20.

Egloff, Daniel, Markus Leippold, and Liuren Wu, 2010, The term structure of variance swap rates and optimal variance swap investments, *Journal of Financial and Quantitative Analysis* 45, 1279–1310.

Fengler, Matthias R., 2005, *Semiparametric Modeling of Implied Volatility*. (Springer, Berlin).

Foresi, Silverio, and Liuren Wu, 2005, Crash-o-phobia: A domestic fear or a worldwide concern? *Journal of Derivatives* 13, 8–21.

Gershon, David, 2018, Model independent multi-asset volatility smile with empirical confirmation, Working paper, Hebrew University of Jerusalem.

Giglio, Stefano, and Bryan T. Kelly, 2018, Excess volatility: Beyond discount rates, *Quarterly Journal of Economics* 133, 71–127.

Hafner, Reinhold, 2004, *Stochastic Implied Volatility: A Factor-Based Model*. (Springer, Berlin).

Heath, David, Robert Jarrow, and Andrew Morton, 1992, Bond pricing and the term structure of interest rates: A new technology for contingent claims valuation, *Econometrica* 60, 77–105.

Heston, Steven L., 1993, Closed-form solution for options with stochastic volatility, with application to bond and currency options, *Review of Financial Studies* 6, 327–343.
Hodges, Hardy M., 1996, Arbitrage bounds of the implied volatility strike and term structures of European-style options, *Journal of Derivatives* 3, 23–32.
Hu, Guanglian, and Kris Jacobs, 2017, Volatility and expected option returns, Working paper, University of Houston.
Hull, John C., 2009, *Options, Futures, and Other Derivatives* (Pearson Prentice Hall, New York).
Israelov, Roni, and Bryan T. Kelly, 2017, Forecasting the distribution of option returns, Working paper, AQR and University of Chicago.
Jones, Christopher S., 2006, A nonlinear factor analysis of S&P 500 index option returns, *Journal of Finance* 61, 2325–2363.
Merton, Robert C., 1973, Theory of rational option pricing, *Bell Journal of Economics and Management Science* 4, 141–183.
Newey, Whitney K., and Kenneth D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703–708.
Schönbucher, Philipp J., 1999, A market model for stochastic implied volatility, *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* 357, 2071–2092.
Shue, Kelly, and Tichard R. Townsend, 2018, Can the market multiply and divide? non-proportional thinking in financial markets, Working paper, Yale University and UC San Diego.
Wu, Liuren, 2006, Dampened power law: Reconciling the tail behavior of financial asset returns, *Journal of Business* 79, 1445–1474.
Wu, Liuren, and Jingyi Zhu, 2016, Simple robust hedging with nearby contracts, *Journal of Financial Econometrics* 15, 1–35.
Wystup, Uwe, 2010, Vanna-volga pricing, in Rama Cont, ed.: *Encyclopedia of Quantitative Finance* (John Wiley & Sons, Chichester).

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Replication code.