Demand-Based Option Pricing

Nicolae Gârleanu
University of California at Berkeley, CEPR, and NBER

Lasse Heje Pedersen
New York University, CEPR, and NBER

Allen M. Poteshman
University of Illinois at Urbana-Champaign

We model demand-pressure effects on option prices. The model shows that demand pressure in one option contract increases its price by an amount proportional to the variance of the unhedgeable part of the option. Similarly, the demand pressure increases the price of any other option by an amount proportional to the covariance of the unhedgeable parts of the two options. Empirically, we identify aggregate positions of dealers and end-users using a unique dataset, and show that demand-pressure effects make a contribution to well-known option-pricing puzzles. Indeed, time-series tests show that demand helps explain the overall expensiveness and skew patterns of index options, and cross-sectional tests show that demand impacts the expensiveness of single-stock options as well.

One of the major achievements of financial economics is the no-arbitrage theory that determines derivative prices independently of investor demand. Building on the seminal contributions of Black and Scholes (1973) and Merton (1973), a large literature develops various parametric implementations of the theory. This literature is surveyed by Bates (2003), who emphasizes that it cannot fully capture—much less explain—the empirical properties of option prices and concludes that there is a need for a new approach to pricing derivatives that focuses on the "financial intermediation of the underlying risks by option market-makers" (Bates 2003, p. 400).

We are grateful for helpful comments from David Bates, Nick Bollen, Oleg Bondarenko, Menachem Brenner, Andrea Buraschi, Josh Coval, Domenico Cuoco, Apoorva Koticha, Sophie Ni, Jun Pan, Neil Pearson, Josh White, and especially from Steve Figlewski, as well as from seminar participants at the Caesarea Center Conference, Columbia University, Cornell University, Dartmouth University, Harvard University, HEC Lausanne, the 2005 Inquire Europe Conference, London Business School, the 2005 NBER Behavioral Finance Conference, the 2005 NBER Universities Research Conference, the NYU-ISE Conference on the Transformation of Option Trading, New York University, Nomura Securities, Oxford University, Stockholm Institute of Financial Research, Texas A&M University, University of California at Berkeley, University of Chicago, University of Illinois at Urbana-Champaign, University of Pennsylvania, University of Vienna, and the 2005 WFA Meetings. Send correspondence to Nicolae Gârleanu, Haas School of Business, University of California, Berkeley, CA 94720-1900. E-mail: garleanu@haas.berkeley.edu.

doi:10.1093/rfs/hhp005 Advance Access publication February 25, 2009

We take on this challenge. Our model departs fundamentally from the no-arbitrage framework by recognizing that option market-makers cannot perfectly hedge their inventories, and, consequently, option demand impacts option prices. We obtain explicit expressions for the effects of demand on option prices, provide empirical evidence consistent with the demand-pressure model using a unique dataset, and show that demand-pressure effects can play a role in resolving the main option-pricing puzzles.

The starting point of our analysis is that options are traded because they are useful and, therefore, options cannot be redundant for all investors (Hakansson 1979). We denote the agents who have a fundamental need for option exposure as "end-users."

Intermediaries such as market-makers provide liquidity to end-users by taking the other side of the end-user net demand. If competitive intermediaries can hedge perfectly—as in a Black-Scholes-Merton economy—then option prices are determined by no-arbitrage and demand pressure has no effect. In reality, however, even intermediaries cannot hedge options perfectly—that is, even they face incomplete markets—because of the impossibility of trading continuously, stochastic volatility, jumps in the underlying, and transaction costs (Figlewski 1989).[1] In addition, intermediaries are sensitive to risk, e.g., because of capital constraints and agency.

In light of these facts, we consider how options are priced by competitive risk-averse dealers who cannot hedge perfectly. In our model, dealers trade an arbitrary number of option contracts on the same underlying at discrete times. Since the dealers trade many option contracts, certain risks net out, while others do not. The dealers can hedge part of the remaining risk of their derivative positions by trading the underlying security and risk-free bonds. We consider a general class of distributions for the underlying, which can accommodate stochastic volatility and jumps. Dealers trade options with end-users. The model is agnostic about the end-users' reasons for trade, which are irrelevant for our results and their empirical implementation.

We compute equilibrium prices as functions of demand pressure, that is, the prices that induce the utility-maximizing dealers to supply precisely the option quantities that the end-users demand. We show explicitly how demand pressure enters into the pricing kernel. Intuitively, a positive demand pressure in an option increases the pricing kernel in the states of nature in which an optimally hedged position has a positive payoff. This pricing-kernel effect increases the price of the option, which entices the dealers to sell it. Specifically, a marginal change in the demand pressure in an option contract increases its price by an amount proportional to the variance of the unhedgeable part of the option, where the variance is computed under a certain probability measure depending on the demand. Similarly, demand pressure increases the price of any

[1] Options may also be impossible to replicate due to asymmetric information (Back 1993; Easley, O'Hara, and Srinivas 1998).

other option by an amount proportional to the covariance of their unhedgeable parts. Hence, while demand pressure in a particular option raises its price, it also raises the prices of other options on the same underlying. Our main theoretical results relating option-price effects to the variance or covariance of the unhedgeable part of option-price changes hold regardless of the source of market incompleteness. The magnitudes of the variances and covariances, and hence of the demand-based option-price effects, depend upon the particular source of market incompleteness. Empirically, we test the specific predictions of the model under the assumptions that market incompleteness stems from discrete trading, stochastic volatility, or jumps.

We use a unique dataset to identify aggregate daily positions of dealers and end-users. In particular, we define dealers as market-makers and end-users as proprietary traders and customers of brokers.[2] We are the first to document that end-users have a net long position in S&P 500 index options with large net positions in out-of-the-money (OTM) puts.[3] Since options are in zero net supply, this implies that dealers are short index options.[4] We estimate that these large short dealer positions lead to daily delta-hedged profits and losses varying between \$100 million and \$-100 million, and cumulative dealer profits of approximately \$800 million over our six-year sample. Hence, consistent with our framework, dealers face significant unhedgeable risk and are compensated for bearing it. Introducing entry of dealers into our model, we show that the total risk-bearing market-making capacity increases with the demand pressure, and we estimate empirically that the market-maker's risk-return profile is consistent with equilibrium entry.

The end-user demand for index options can help to explain the two puzzles that index options appear to be expensive, and that low-moneyness options seem to be especially expensive (Rubinstein 1994; Longstaff 1995; Bates 2000; Jackwerth 2000; Coval and Shumway 2001; Bondarenko 2003; Amin, Coval, and Seyhun 2004; Driessen and Maenhout 2008). In the time series, the model-based impact of demand for index options is positively related to their expensiveness, measured by the difference between their implied volatility and the volatility measure of Bates (2006). Indeed, we estimate that on the order of one-third of index-option expensiveness can be accounted for by demand effects.[5] In addition, the link between demand and prices is stronger following recent dealer losses, as would be expected if dealers are more risk averse at such times. Likewise, the steepness of the smirk, measured by the difference

[2] The empirical results are robust to classifying proprietary traders as either dealers or end-users.

[3] These positions are consistent with the end-users suffering from "crashophobia," as suggested by Rubinstein (1994).

[4] Option traders recognize the importance of demand effects; e.g., Vanessa Gray, director of global equity derivatives, Dresdner Kleinwort Benson, states that option-implied volatility skew "is heavily influenced by supply and demand factors."

[5] Premiums for stochastic volatility or jump risk, as well as premiums for other risk factors, in all likelihood are also contributing to option expensiveness.

between the implied volatilities of low-moneyness options and at-the-money (ATM) options, is positively related to the skew of option demand.

Another option-pricing puzzle is the significant difference between index-option prices and the prices of single-stock options, despite the relative similarity of the underlying distributions (e.g., Bakshi, Kapadia, and Madan 2003; Bollen and Whaley 2004). In particular, single-stock options appear cheaper and their smile is flatter. Consistently, we find that the demand pattern for single-stock options is very different from that of index options. For instance, end-users are net short single-stock options—not long, as in the case of index options. Demand patterns further help to explain the cross-sectional pricing of single-stock options. Indeed, individual stock options are relatively cheaper for stocks with more negative demand for options.

The article is related to several strands of literature. First, a large literature documents option puzzles relative to the existing models (cited above).[6]

Second, the literature on option pricing in the context of trading frictions and incomplete markets derives bounds on option prices (Soner, Shreve, and Cvitanic 1995; Bernardo and Ledoit 2000; Cochrane and Saa-Requejo 2000; Constantinides and Perrakis 2002; Constantinides, Jackwerth, and Perrakis forthcoming). Rather than deriving bounds, we compute explicit prices based on the demand pressure by end-users. We further complement this literature by taking portfolio considerations into account, that is, the effect of demand for one option on the prices of other options.

Third, the general idea of demand pressure effects goes back at least to Keynes (1923) and Hicks (1939), who considered futures markets. Our model is the first to apply this idea to option pricing and to incorporate the important features of options markets, namely dynamic trading of many assets, hedging using the underlying and bonds, stochastic volatility, and jumps. The generality of our model also makes it applicable to other situations—e.g., stock-index additions (Shleifer 1986, Wurgler and Zhuravskaya 2002, Greenwood 2005), the fixed-income market (Newman and Rierson 2004), mortgage-backed security markets (Gabaix, Krishnamurthy, and Vigneron 2007), futures markets (de Roon, Nijman, and Veld 2000), and option end-users' over-reaction to changes in implied volatility (Stein 1989; Poteshman 2001).

Fourth, the literature on utility-based option pricing derives the price of the first "marginal" option that would make an agent indifferent between buying the option and not buying it (Rubinstein 1976; Brennan 1979; Stapleton and Subrahmanyam 1984; Hugonnier, Kramkov, and Schachermayer 2005, and references therein), and we show how option-prices change when demand is nontrivial.

A closely related paper is Bollen and Whaley (2004), which demonstrates that changes in implied volatility are correlated with signed option volume. These

[6] In contrast to these papers, Benzoni, Collin-Dufresne, and Goldstein (2005) find that option prices can be rationalized by certain preferences combined with persistent long-run jump risk.

empirical results set the stage for our analysis by showing that *changes* in option demand lead to *changes* in option prices while leaving open the question of whether the *level* of option demand impacts the overall *level* (i.e., expensiveness) of option prices or the overall shape of implied-volatility curves.[7] We complement Bollen and Whaley (2004) by providing a theoretical model, by investigating empirically the relationship between the level of end-user demand for options and the level and overall shape of implied-volatility curves, and by testing precise quantitative implications of our model. In particular, we document that end-users tend to have a net long SPX option position and a short equity-option position, thus helping to explain the relative expensiveness of index options. We also show that there is a strong downward skew in the net demand of index but not equity options, which helps to explain the difference in the shapes of their overall implied-volatility curves. In addition, we demonstrate that option prices are better explained by model-based rather than simple nonmodel-based use of demand.

1. A Model of Demand Pressure

We consider a discrete-time infinite-horizon economy. There exist a risk-free asset paying interest at the rate of $R_f - 1$ per period and a risky security that we refer to as the "underlying" security. At time t, the underlying has an exogenous strictly positive price[8] of S_t, dividend D_t, and an excess return of $R_t^e = (S_t + D_t)/S_{t-1} - R_f$. The distribution of future prices and returns is characterized by a Markov state variable X_t, which includes the current underlying price level, $X_t^1 = S_t$, and may also include the current level of volatility, the current jump intensity, etc. We assume that (R_t^e, X_t) satisfies a Feller-type condition (made precise in the Appendix) and that X_t is bounded for every t.

The economy further has a number of "derivative" securities, whose prices are to be determined endogenously. A derivative security is characterized by its index $i \in I$, where i collects the information that identifies the derivative and its payoffs. For a European option, for instance, the strike price, maturity date, and whether the option is a "call" or "put" suffice. The set of derivatives traded at time t is denoted by I_t, and the vector of prices of traded securities is $p_t = (p_t^i)_{i \in I_t}$.

The payoffs of the derivatives depend on X_t. We note that the theory is completely general and does not require that the "derivatives" have payoffs that depend on the underlying price. In principle, the derivatives could be any securities—in particular, any securities whose prices are affected by demand

[7] Indeed, Bollen and Whaley (2004) find that a nontrivial part of the option-price impact from day t signed option volume dissipates by day $t + 1$.

[8] All random variables are defined on a probability space $(\Omega, \mathcal{F}, Pr)$, with an associated filtration $\{\mathcal{F}_t : t \geq 0\}$ of sub-σ-algebras representing the resolution over time of information commonly available to agents.

pressure. Further, it is straightforward to extend our results to a model with any number of exogenously priced securities, as discussed in Section 2.4. While we use the model to study options in particular, we think that the generality helps to illuminate the driving forces behind the results and, further, it allows future applications of the theory in other markets.

The economy is populated by two kinds of agents: "dealers" and "end-users." Dealers are competitive and there exists a representative dealer who has constant absolute-risk aversion, that is, his utility for remaining life time consumption is

$$U(C_t, C_{t+1}, \ldots) = \mathrm{E}_t\left[\sum_{v=t}^{\infty} \rho^{v-t} u(C_v)\right], \tag{1}$$

where $u(c) = -\frac{1}{\gamma}e^{-\gamma c}$ and $\rho < 1$ is a discount factor. At any time t, the dealer must choose the consumption C_t, the dollar investment in the underlying θ_t, and the number of derivatives held $q_t = (q_t^i)_{i \in I_t}$, so as to maximize his utility while satisfying the transversality condition $\lim_{t\to\infty} \mathrm{E}[\rho^{-t} e^{-kW_t}] = 0$. The dealer's wealth evolves as

$$W_{t+1} \;=\; (W_t - C_t)R_f + q_t(p_{t+1} - R_f p_t) + \theta_t R_{t+1}^e. \tag{2}$$

In the real world, end-users trade options for a variety of reasons, such as portfolio insurance, agency reasons, behavioral reasons, institutional reasons, etc. Rather than trying to capture these various trading motives endogenously, we assume that end-users have an exogenous aggregate demand for derivatives of $d_t = (d_t^i)_{i \in I_t}$ at time t. The distribution of future demand is characterized by X_t. We also assume, for technical reasons, that demand pressure is zero after some time $\overline{T}$, that is, $d_t = 0$ for $t > \overline{T}$.

Derivative prices are set through the interaction between dealers and end-users in a competitive equilibrium.

Definition 1. *A price process $p_t = p_t(d_t, X_t)$ is a (competitive Markov) equilibrium if, given p, the representative dealer optimally chooses a derivative holding q such that derivative markets clear, i.e., $q + d = 0$.*

We note that the assumption of inelastic end-user demand is made for notational simplicity only, and is unimportant for the results we derive below. The key to our asset-pricing approach is the insight that, by observing the aggregate quantities held by dealers in equilibrium, one can determine the derivative prices consistent with the dealers' utility maximization, that is, "invert prices from quantities." Our goal is to determine how derivative prices depend on the demand pressure d coming from end-users. All that matters is that end-users have demand curves that result in dealers choosing to hold, at the market prices, a position of $q = -d$ that we observe in the data.

To determine the representative dealer's optimal behavior, we consider his value function $J(W; t, X)$, which depends on his wealth W, the state of nature X, and time t. Then, the dealer solves the following maximization problem:

$$\max_{C_t, q_t, \theta_t} \quad -\frac{1}{\gamma} e^{-\gamma C_t} + \rho \mathrm{E}_t[J(W_{t+1}; t+1, X_{t+1})] \tag{3}$$

$$\text{s.t.} \quad W_{t+1} = (W_t - C_t)R_f + q_t(p_{t+1} - R_f p_t) + \theta_t R^e_{t+1}. \tag{4}$$

The value function is characterized in the following lemma.

Lemma 1. *If $p_t = p_t(d_t, X_t)$ is the equilibrium price process and $k = \frac{\gamma(R_f - 1)}{R_f}$, then the dealer's value function and optimal consumption are given by*

$$J(W_t; t, X_t) = -\frac{1}{k} e^{-k(W_t + G_t(d_t, X_t))} \tag{5}$$

$$C_t = \frac{R_f - 1}{R_f}(W_t + G_t(d_t, X_t)) \tag{6}$$

and the stock and derivative holdings are characterized by the first-order conditions

$$0 = E_t\left[e^{-k(\theta_t R^e_{t+1} + q_t(p_{t+1} - R_f p_t) + G_{t+1}(d_{t+1}, X_{t+1}))} R^e_{t+1}\right] \tag{7}$$

$$0 = E_t\left[e^{-k(\theta_t R^e_{t+1} + q_t(p_{t+1} - R_f p_t) + G_{t+1}(d_{t+1}, X_{t+1}))}(p_{t+1} - R_f p_t)\right], \tag{8}$$

where, for $t \leq T$, $G_t(d_t, X_t)$ is derived recursively using (7), (8), and

$$e^{-kR_f G_t(d_t, X_t)} = R_f \rho E_t\left[e^{-k\left(q_t(p_{t+1} - R_f p_t) + \theta_t R^e_{t+1} + G_{t+1}(d_{t+1}, X_{t+1})\right)}\right] \tag{9}$$

and for $t > T$, the function $G_t(d_t, X_t) = \bar{G}(X_t)$, where $(\bar{G}(X_t), \bar{\theta}(X_t))$ solves

$$e^{-kR_f \bar{G}(X_t)} = R_f \rho E_t\left[e^{-k\left(\bar{\theta}_t R^e_{t+1} + \bar{G}(X_{t+1})\right)}\right] \tag{10}$$

$$0 = E_t\left[e^{-k\left(\bar{\theta}_t R^e_{t+1} + \bar{G}(X_{t+1})\right)} R^e_{t+1}\right]. \tag{11}$$

The optimal consumption is unique. The optimal security holdings are unique, provided that their payoffs are linearly independent.

While dealers compute optimal positions given prices, we are interested in inverting this mapping and compute the prices that make a given position optimal. The following proposition ensures that this inversion is possible.

Proposition 1. *Given any demand pressure process d for end-users, there exists a unique equilibrium price process p.*

Before considering explicitly the effect of demand pressure, we make a couple of simple "parity" observations that show how to treat derivatives that are linearly dependent, such as European puts and calls with the same strike and maturity. For simplicity, we do this only in the case of a nondividend paying underlying, but the results can easily be extended. We consider two derivatives i and j such that a nontrivial linear combination of their payoffs lies in the span of exogenously priced securities, i.e., the underlying and the bond:

Proposition 2. *Suppose that $D_t = 0$ and $p_T^i = p_T^j + \alpha + \beta S_T$. Then:*
(i) For any demand pressure, d, the equilibrium prices of the two derivatives are related by

$$p_t^i = p_t^j + \alpha R_f^{-(T-t)} + \beta S_t. \tag{12}$$

(ii) Changing the end-user demand from (d_t^i, d_t^j) to $(d_t^i + a, d_t^j - a)$, for any $a \in \mathbb{R}$, has no effect on equilibrium prices.

The first part of the proposition is a general version of the well-known put-call parity. It shows that if payoffs are linearly dependent, then so are prices.

The second part of the proposition shows that linearly dependent derivatives have the same demand-pressure effects on prices. Hence, in our empirical exercise, we can aggregate the demand of calls and puts with the same strike and maturity. That is, a demand pressure of d^i calls and d^j puts is the same as a demand pressure of $d^i + d^j$ calls and 0 puts (or vice versa).

2. Price Effects of Demand Pressure

To see where we are going with the theory, consider the empirical problem that we ultimately face: On any given day, around 120 SPX option contracts of various maturities and strike prices are traded. The demands for all these different options potentially affect the price of, say, the one-month ATM SPX option because all of these options expose the market-makers to unhedgeable risk. What is the aggregate effect of all these demands?

The model answers this question by showing how to compute the impact of demand d_t^j for any one derivative on the price p_t^i of the one-month ATM option. The aggregate effect is then the sum of all of the individual demand effects, that is, the sum of all the demands weighted by their model-implied price impacts $\partial p_t^i / \partial d_t^j$.

We first characterize $\partial p_t^i / \partial d_t^j$ in complete generality, as well as other general demand effects on prices (Section 2.1). We then show how to compute $\partial p_t^i / \partial d_t^j$ specifically when unhedgeable risk arises from, respectively, discrete-time hedging, jumps in the underlying asset price, and stochastic-volatility risk (Section 2.2). Section 2.3 studies the price effect with an endogenous number of dealers, which gives rise to testable implications relevant for our cross-section of equity options. Finally, Section 2.4 generalizes to multiple underlying securities.

2.1 General results

We think of the price p, the hedge position θ_t in the underlying, and the consumption function G as functions of d_t^j and X_t. Alternatively, we can think of the dependent variables as functions of the dealer holding q_t^j and X_t, keeping in mind the equilibrium relation that $q = -d$. For now we use this latter notation.

At maturity date T, an option has a known price p_T. At any prior date t, the price p_t can be found recursively by "inverting" (8) to get

$$p_t = \frac{\mathrm{E}_t\left[e^{-k\left(\theta_t R_{t+1}^e + q_t p_{t+1} + G_{t+1}\right)} p_{t+1}\right]}{R_f \mathrm{E}_t\left[e^{-k\left(\theta_t R_{t+1}^e + q_t p_{t+1} + G_{t+1}\right)}\right]}, \tag{13}$$

where the hedge position in the underlying, θ_t, solves

$$0 = \mathrm{E}_t\left[e^{-k\left(\theta_t R_{t+1}^e + q_t p_{t+1} + G_{t+1}\right)} R_{t+1}^e\right], \tag{14}$$

and where G is computed recursively as described in Lemma 1. Equations (13) and (14) can be written in terms of a demand-based pricing kernel:

Theorem 1. *Prices p and the hedge position θ satisfy*

$$p_t = E_t\left(m_{t+1}^d p_{t+1}\right) = \frac{1}{R_f} E_t^d (p_{t+1}) \tag{15}$$

$$0 = E_t\left(m_{t+1}^d R_{t+1}^e\right) = \frac{1}{R_f} E_t^d \left(R_{t+1}^e\right), \tag{16}$$

where the pricing kernel m^d is a function of demand pressure d:

$$m_{t+1}^d = \frac{e^{-k\left(\theta_t R_{t+1}^e + q_t p_{t+1} + G_{t+1}\right)}}{R_f E_t\left[e^{-k\left(\theta_t R_{t+1}^e + q_t p_{t+1} + G_{t+1}\right)}\right]} \tag{17}$$

$$= \frac{e^{-k\left(\theta_t R_{t+1}^e - d_t p_{t+1} + G_{t+1}\right)}}{R_f E_t\left[e^{-k\left(\theta_t R_{t+1}^e - d_t p_{t+1} + G_{t+1}\right)}\right]}, \tag{18}$$

and E_t^d is expected value with respect to the corresponding risk-neutral measure, i.e., the measure with a Radon-Nikodym derivative with respect to the objective measure of $R_f m_{t+1}^d$.

To understand this pricing kernel, suppose for instance that end-users want to sell derivative i such that $d_t^i < 0$, and that this is the only demand pressure. In equilibrium, dealers take the other side of the trade, buying $q_t^i = -d_t^i > 0$

units of this derivative, while hedging their derivative holding using a position θ_t in the underlying. The pricing kernel is small whenever the "unhedgeable" part $q_t p_{t+1} + \theta_t R^e_{t+1}$ is large. Hence, the pricing kernel assigns a low value to states of nature in which a hedged position in the derivative pays off profitably, and it assigns a high value to states in which a hedged position in the derivative has a negative payoff. This pricing-kernel effect decreases the price of this derivative, which is what entices the dealers to buy it.

It is interesting to consider the first-order effect of demand pressure on prices. In order to do so, we first define the unhedgeable part of the price changes of a security.

Definition 2. *The unhedgeable price change $\bar{p}^k_{t+1}$ of any security k is defined as its excess return $p^k_{t+1} - R_f p^k_t$ optimally hedged with the stock position $\frac{Cov^d_t(p^k_{t+1}, R^e_{t+1})}{Var^d_t(R^e_{t+1})}$:*

$$\bar{p}^k_{t+1} = R_f^{-1}\left(p^k_{t+1} - R_f p^k_t - \frac{Cov^d_t\left(p^k_{t+1}, R^e_{t+1}\right)}{Var^d_t\left(R^e_{t+1}\right)} R^e_{t+1}\right). \tag{19}$$

We prove in the Appendix the following result.

Theorem 2. *The sensitivity of the price of security i to demand pressure in security j is proportional to the covariance of their unhedgeable risks:*

$$\frac{\partial p^i_t}{\partial d^j_t} = \gamma(R_f - 1)E^d_t\left(\bar{p}^i_{t+1}\bar{p}^j_{t+1}\right) = \gamma(R_f - 1)Cov^d_t\left(\bar{p}^i_{t+1}, \bar{p}^j_{t+1}\right). \tag{20}$$

This result is intuitive: it states that the demand pressure in an option j increases the option's own price by an amount proportional to the variance of the unhedgeable part of the option and the aggregate risk aversion of dealers. We note that since a variance is always positive, the demand-pressure effect on the security itself is naturally always positive. Further, this demand pressure affects another option i by an amount proportional to the covariance of their unhedgeable parts. Under the condition stated below, we can show that this covariance is positive, and therefore that demand pressure in one option also increases the price of other options on the same underlying.

Proposition 3. *Demand pressure in any security j:*
(i) increases its own price, that is, $\frac{\partial p^j_t}{\partial d^j_t} \geq 0$;
(ii) increases the price of another security i, that is, $\frac{\partial p^i_t}{\partial d^j_t} \geq 0$, provided that $E^d_t[p^i_{t+1} | S_{t+1}]$ and $E^d_t[p^j_{t+1} | S_{t+1}]$ are convex functions of S_{t+1} and $Cov^d_t(p^i_{t+1}, p^j_{t+1} | S_{t+1}) \geq 0$.

The conditions imposed in part (ii) are natural. First, we require that prices inherit the convexity property of the option payoffs in the underlying price. Convexity lies at the heart of this result, which, informally speaking, states that higher demand for convexity (or gamma, in option-trader lingo) increases its price, and therefore those of all options. Second, we require that $\mathrm{Cov}_t^d(p_{t+1}^i, p_{t+1}^j \mid S_{t+1}) \geq 0$, that is, changes in the other variables have a similar impact on both option prices—for instance, both prices are increasing in the volatility or demand level. Note that both conditions hold if both options mature after one period. The second condition also holds if option prices are homogenous (of degree 1) in (S, K), where K is the strike, and S_t is independent of $X_t^{-1} \equiv (X_t^2, \ldots, X_t^n)$.

It is interesting to consider the total price that end-users pay for their demand d_t at time t. Vectorizing the derivatives from Theorem 2, we can first-order approximate the price around zero demand as

$$p_t \approx p_t(d_t = 0) + \gamma(R_f - 1)\mathrm{E}_t^d(\bar{p}_{t+1}\bar{p}_{t+1}')d_t. \tag{21}$$

Hence, the total price paid for the d_t derivatives is

$$d_t' p_t = d_t' p_t(d_t = 0) + \gamma(R_f - 1)d_t'\mathrm{E}_t^d(\bar{p}_{t+1}\bar{p}_{t+1}')d_t \tag{22}$$

$$= d_t' p_t(d_t = 0) + \gamma(R_f - 1)\mathrm{Var}_t^d(d_t'\bar{p}_{t+1}). \tag{23}$$

The first term $d_t' p_t(d_t = 0)$ is the price that end-users would pay if their demand pressure did not affect prices. The second term is total variance of the unhedgeable part of all of the end-users' positions.

While Proposition 3 shows that demand for an option increases the prices of all options, the size of the price effect is, of course, not the same for all options. Nor is the effect on implied volatilities the same. Under certain conditions, demand pressure in low-strike options has a larger impact on the implied volatility of low-strike options, and conversely for high-strike options. The following proposition makes this intuitively appealing result precise. For simplicity, the proposition relies on unnecessarily restrictive assumptions. We let $p(p, K, d)$, respectively $p(c, K, d)$, denote the price of a put, respectively a call, with strike price K and one period to maturity, where d is the demand pressure. It is natural to compare low-strike and high-strike options that are "equally far out of the money." We do this by considering an OTM put with the same price as an OTM call.

Proposition 4. *Assume that the one-period risk-neutral distribution of the underlying return is symmetric and that X_{t+1}^{-1} is independent of S_{t+1}. Consider demand pressure $d_{t>0}$ in an option with strike $K < R_f S_t$ that matures after one trading period. Then there exists a value $\bar{K}$ such that, for all $K' \leq \bar{K}$ and K'' such that $p(p, K', 0) = p(c, K'', 0)$, it holds that $p(p, K', d) > p(c, K'', d)$. That is, the price of the OTM put $p(p, K', \cdot)$ is more affected by the demand* *pressure than the price of OTM call $p(c, K'', \cdot)$. The reverse conclusion applies if there is demand for a high-strike option.*

Future demand pressure in a derivative j also affects the current price of derivative i. As above, we consider the first-order price effect. This is slightly more complicated, however, since we cannot differentiate with respect to the unknown future demand pressure. Instead, we "scale" the future demand pressure, that is, we consider future demand pressures $\tilde{d}_s^j = \epsilon d_s^j$ for fixed d (equivalently, $\tilde{q}_s^j = \epsilon q_s^j$) for some $\epsilon \in \mathbb{R}$, $\forall s > t$, and $\forall j$.

Theorem 3. *Let $p_t(0)$ denote the equilibrium derivative prices with 0 demand pressure. Fixing a process d with $d_t = 0$ for all $t > T$ and a given T, the equilibrium prices p with a demand pressure of ϵd is*

$$p_t = p_t(0) + \gamma(R_f - 1)\left[E_t^0\big(\bar{p}_{t+1}\bar{p}_{t+1}'\big)d_t + \sum_{s>t} R_f^{-(s-t)} E_t^0\big(\bar{p}_{s+1}\bar{p}_{s+1}' d_s\big)\right]\epsilon + O(\epsilon^2). \tag{24}$$

This theorem shows that the impact of current demand pressure d_t on the price of a derivative i is given by the amount of hedging risk that a marginal position in security i would add to the dealer's portfolio, that is, it is the sum of the covariances of its unhedgeable part with the unhedgeable part of all the other securities, multiplied by their respective demand pressures. Further, the impact of future demand pressures d_s is given by the expected future hedging risks. Of course, the impact increases with the dealers' risk aversion.

Next, we specialize the setup to several different sources of unhedgeable risk to show how to compute these covariances, and therefore the price impacts, explicitly.

2.2 Implementation: specific cases

We consider now three examples of unhedgeable risk for the dealers, arising from (i) the inability to hedge continuously, (ii) jumps in the underlying price, and (iii) stochastic-volatility risk, respectively. We focus on small hedging periods Δ_t and derive the results informally while relegating a more rigorous treatment to the Appendix. The continuously compounded risk-free interest rate is denoted by r, i.e., the risk-free return over one Δ_t time period is $R_f = e^{r\Delta_t}$.

We are interested in the price $p_t^i = p_t^i(d_t, X_t)$ of option i as a function of demand pressure d_t and the state variable X_t. (Remember that $S_t = X_t^1$.) We denote the option price without demand pressure by f, that is, $f^i(t, X_t) := p_t^i(d_t = 0, X_t)$, and assume throughout that f is smooth for $t < T$. We use the notation $f^i = f^i(t, X_t)$, $f_t^i = \frac{\partial}{\partial t} f^i(t, X_t)$, $f_S^i = \frac{\partial}{\partial S} f^i(t, X_t)$, $f_{SS}^i = \frac{\partial^2}{\partial S^2} f^i(t, X_t)$, $\Delta S = S_{t+1} - S_t$, and so on.

Case 1: Discrete-time trading. To focus on the specific risk due to discrete-time trading (rather than continuous trading), we consider a stock price that is a diffusion process driven by a Brownian motion[9] with no other state variables (i.e., $X = S$). In this case, markets would be complete with continuous trading, and, hence, the dealer's hedging risk arises solely from his trading only at discrete times, spaced Δ_t time units apart.

The change in the option price evolves approximately according to

$$p^i_{t+1} \cong f^i + f^i_S \Delta S + \frac{1}{2} f^i_{SS} (\Delta S)^2 + f^i_t \Delta_t, \tag{25}$$

while the unhedgeable option-price change is

$$e^{r\Delta_t} \bar{p}^i_{t+1} = p^i_{t+1} - e^{r\Delta_t} p^i_t - f^i_S \big(S_{t+1} - e^{r\Delta_t} S_t\big) \tag{26}$$

$$\cong -r\Delta_t f^i + f^i_t \Delta_t + r\Delta_t f^i_S S_t + \frac{1}{2} f^i_{SS} (\Delta S)^2. \tag{27}$$

The covariance of the unhedgeable parts of two options i and j is

$$\mathrm{Cov}_t\big(e^{r\Delta_t} \bar{p}^i_{t+1}\,,\ e^{r\Delta_t} \bar{p}^j_{t+1}\big) \cong \frac{1}{4} f^i_{SS} f^j_{SS} Var_t((\Delta S)^2), \tag{28}$$

so that, by Theorem 2, we conclude that the effect on the price of demand at $d = 0$ is

$$\frac{\partial p^i_t}{\partial d^j_t} = \frac{\gamma r Var_t((\Delta S)^2)}{4} f^i_{SS} f^j_{SS} + o\big(\Delta^2_t\big) \tag{29}$$

and the effect on the Black-Scholes implied volatility $\hat{\sigma}^i_t$ is

$$\frac{\partial \hat{\sigma}^i_t}{\partial d^j_t} = \frac{\gamma r Var_t((\Delta S)^2)}{4} \frac{f^i_{SS}}{\nu^i} f^j_{SS} + o\big(\Delta^2_t\big), \tag{30}$$

where ν^i is the Black-Scholes vega.[10] Interestingly, the ratio of the Black-Scholes gamma to the Black-Scholes vega, f^i_{SS}/ν^i, does not depend on moneyness, so the first-order effect of demand with discrete trading risk is to change the level, but not the slope, of the implied-volatility curves.

Intuitively, the impact of the demand for options of type j depends on the gamma of these options, f^j_{SS}, since the dealers cannot hedge the nonlinearity of the payoff.

The calculations above show that the effect of discrete-time trading is small if hedging is frequent. More precisely, the effect is of the order of $Var_t((\Delta S)^2)$,

[9] Strictly speaking, we need all price processes to be bounded, e.g., truncated.

[10] Even though the volatility is constant within the Black-Scholes model, we follow the standard convention that defines the Black-Scholes implied volatility as the volatility that, when fed into the Black-Scholes model, makes the model price equal to the option price, and the Black-Scholes vega as the partial derivative measuring the change in the option price when the volatility fed into the Black-Scholes model changes.

namely Δ_t^2. Hence, adding up T/Δ_t terms of this magnitude—corresponding to demand in each period between time 0 and maturity T—results in a total effect of order Δ_t, which approaches zero as Δ_t approaches zero. This is consistent with the Black-Scholes-Merton result of perfect hedging in continuous time. As the next examples show, the risks of jumps and stochastic volatility do not vanish for small Δ_t (specifically, they are of order Δ_t).

Case 2: Jumps in the underlying. Suppose now that S is a jump diffusion with i.i.d. jump size and jump intensity π (i.e., jump probability over a period of approximately $\pi\Delta_t$).

The unhedgeable price change is

$$e^{r\Delta_t}\bar{p}^i_{t+1} \cong -r\Delta_t f^i + f^i_t\Delta_t + r\Delta_t f^i_S S_t + \left(f^i_S S_t - \theta^i\right)\Delta S 1_{(\text{no jump})} + \kappa^i 1_{(\text{jump})}, \tag{31}$$

where

$$\kappa^i = f^i(S_t+\eta) - f^i - \theta^i\eta \tag{32}$$

is the unhedgeable risk in case of a jump of size η.

It then follows that the effect on the price of demand at $d = 0$ is

$$\frac{\partial p^i_t}{\partial d^j_t} = \gamma r\left[\left(f^i_S S_t - \theta^i\right)\left(f^j_S S_t - \theta^j\right)\mathrm{Var}_t(\Delta S) + \pi\Delta_t \mathrm{E}_t\left(\kappa^i\kappa^j\right)\right] + o(\Delta_t) \tag{33}$$

and the effect on the Black-Scholes implied volatility $\hat{\sigma}^i_t$ is

$$\frac{\partial \hat{\sigma}^i_t}{\partial d^j_t} = \frac{\gamma r\left[\left(f^i_S S_t - \theta^i\right)\left(f^j_S S_t - \theta^j\right)\mathrm{Var}_t(\Delta S) + \pi\Delta_t \mathrm{E}_t\left(\kappa^i\kappa^j\right)\right]}{\nu^i} + o(\Delta_t). \tag{34}$$

The terms of the form $f^i_S S_t - \theta^i$ arise because the optimal hedge θ differs from the optimal hedge without jumps, $f^i_S S_t$, which means that some of the local noise is being hedged imperfectly. If the jump probability is small, however, then this effect is small (i.e., it is second order in π). In this case, the main effect comes from the jump risk κ (kappa). We note that, while conventional wisdom holds that Black-Scholes gamma is a measure of "jump risk," this is true only for the small local jumps considered in Case 1. Large jumps have qualitatively different implications captured by kappa. For instance, a far-OTM put may have little gamma risk, but, if a large jump can bring the option in the money, the option may have kappa risk. It can be shown that this jump-risk effect (34) means that demand can affect the slope of the implied-volatility curve to the first order and generate a smile.[11]

[11] Of course, the jump risk also generates smiles without demand-pressure effects; the result is that demand can exacerbate these.

Another important source of unhedgeability, itself an important statistical property of underlying prices and therefore playing a significant role in modern option-pricing models, is stochastic volatility. We illustrate below how to calculate the price impact of demand in its presence.

Case 3: Stochastic-volatility risk. We now let the state variable be $X_t = (S_t, \sigma_t)$, where the stock price S is a diffusion with volatility σ_t, which is also a diffusion, driven by an independent Brownian motion. The option price $p_t^i = f^i(t, S_t, \sigma_t)$ has unhedgeable risk given by

$$e^{r\Delta_t}\bar{p}_{t+1}^i = p_{t+1}^i - e^{r\Delta_t} p_t^i - \theta^i R_{t+1}^e \tag{35}$$

$$\cong -r\Delta_t f^i + f_t^i \Delta_t + f_S^i S_t r\Delta_t + f_\sigma^i \Delta\sigma_{t+1}, \tag{36}$$

so that the effect on the price of demand at $d = 0$ is

$$\frac{\partial p_t^i}{\partial d_t^j} = \gamma r \mathrm{Var}(\Delta\sigma) f_\sigma^i f_\sigma^j + o(\Delta_t) \tag{37}$$

and the effect on the Black-Scholes implied volatility $\hat{\sigma}_t^i$ is

$$\frac{\partial \hat{\sigma}_t^i}{\partial d_t^j} = \gamma r \mathrm{Var}(\Delta\sigma) \frac{f_\sigma^i}{\nu^i} f_\sigma^j + o(\Delta_t). \tag{38}$$

Intuitively, volatility risk is captured to the first order by f_σ. This derivative is not exactly the same as Black-Scholes vega, since vega is the price sensitivity to a permanent volatility change, whereas f_σ measures the price sensitivity to a volatility change that may decay. If volatility mean reverts at the rate ϕ, then, for an option with maturity at time $t + T$, we have

$$f_\sigma^i \cong \nu^i \frac{\partial}{\partial \sigma_t} E\left(\frac{\int_t^{t+T} \sigma_s ds}{T} \,\Big|\, \sigma_t\right) = \nu^i \frac{1 - e^{-\phi T}}{\phi T}. \tag{39}$$

Hence, combining (39) with (38) shows that stochastic-volatility risk affects the level, but not the slope, of the implied-volatility curves to the first order.

We make use of each of these three explicitly modeled sources of unhedgeable risk in our empirical work, where we base the model-implied empirical measures of demand impact on the formulae (30), (34), and (38), respectively. Our results could be generalized by introducing a time-varying jump intensity, jumps in the volatility (mathematically, this would be similar to our analysis of jumps in the underlying), or a more complicated correlation structure for the state variables. While such generalizations would add realism, we want to test the effect of the demand in the presence of the most basic sources of unhedgeable risk considered here.

2.3 Equilibrium number of dealers

The number of dealers and their aggregate risk-bearing capacity are determined in equilibrium by dealers' tradeoff between the costs and benefits of making markets. This section shows that the aggregate dealer risk aversion γ can be determined as the outcome of equilibrium dealer entry and, further, provides some natural properties of γ.

We consider an infinitesimal agent with risk aversion γ', who could become a dealer at time $t = 0$ at a cost of M dollars. There is a continuum of available dealers indexed by $i \in [0, \infty)$ with risk aversion $\gamma'(i)$ increasing in i. The distribution of i has no atoms, and is denoted by μ. We assume that $\int_0^\infty \gamma'(i)^{-1} d\mu(i) = \infty$.

Paying the cost M allows the dealer to trade derivatives at all times. This cost could correspond to the cost of a seat on the CBOE, the salaries of traders, the cost of running a back office, etc. In Section 3.3, we estimate the costs and benefits of being a market-maker to be of the same magnitude, consistent with this equilibrium condition.

In the Appendix we show that there exists an equilibrium to the dealer entry game and that, naturally, the least risk-averse dealers $i \in [0, \bar{i}]$ for some $\bar{i} \in \mathbb{R}$ enter the market to profit from price responses to end-user demand. Further, the aggregate dealer demand is the same as that of a representative dealer with risk aversion γ given by $\gamma^{-1} = \int_0^{\bar{i}} \gamma'(i)^{-1} d\mu(i)$.

Proposition 5. *Equilibrium entry of dealers at time 0 implies the following:*

1. *Suppose that the end-user demand is $d_t = \epsilon \bar{d}_t$ for some demand process $\bar{d}$.*
   a. *A higher expected end-user demand leads to more entry of dealers. Specifically, the equilibrium number of dealers increases in ϵ and the equilibrium dealer risk aversion γ decreases in ϵ.*
   b. *If potential dealers have different risk aversions, then prices are more distorted by demand if demand is larger. Rigorously, the absolute price deviation of any unhedgeable option from its zero-demand value increases strictly in ϵ on $[0, \bar{\epsilon}]$, for some $\bar{\epsilon} > 0$.*
   c. *If all potential dealers have the same risk aversion, then derivative prices are independent of ϵ. Nevertheless, derivative prices vary with demand in the time series.*
2. *The equilibrium number of dealers decreases with the cost M of being a dealer. Hence, the aggregate dealer risk aversion γ increases with the (opportunity) cost M.*

Part 1(b) of the proposition states the natural result that increased demand leads to larger price deviations. Indeed, while larger demand leads to entry of dealers, these dealers are increasingly risk averse, leading to the increased demand effect.

Part 1(c) gives the surprising result that the overall level of demand does not affect option prices when all dealers have the same risk aversion because of the entry of dealers. Note, however, that—even in this case—demand affects prices in the time series, that is, at times with more demand, prices are more affected.

This time-series effect would be reduced if dealers could enter at any time. The time 0 entry captures that the decision to set up a trading capability is made only rarely due to significant fixed costs, although, of course, entry and exit does happen over time in the real world.

2.4 Multiple underlying securities

So far, we have considered dealers who trade options on the same underlying, but our results extend to the case in which dealers trade options on multiple underlying assets: Indeed, Theorems 1–2 and Propositions 1–2 continue to hold, subject to treating R^e and θ as vectors and therefore the variances $\mathrm{Var}_t^d(R^e_{t+1})$ and covariances $\mathrm{Cov}_t^d(p^j_{t+1}, R^e_{t+1})$ as matrices. Hence, demand still affects prices through the unhedgeable risk, but now the unhedgeable part is the residual risk after hedging with the multiple underlying securities.

Naturally, demand for an option still increases its own price (part (i) of Proposition 3), and increases the price of other options on that underlying under certain assumptions (as in part (ii) of Proposition 3). The effect of demand for, say, an IBM option on the price of a Microsoft option is, however, more subtle. We want to determine this cross-effect when unhedgeable risk is driven by either (i) discrete-time trading, or (ii) stochastic volatility (defined as in Section 2.2).

For case (i), we suppose that S_1 and S_2 are geometric Brownian motions with instantaneous correlation ρ and compute the covariance of the unhedgeable parts to be

$$
\begin{aligned}
\frac{\partial p_t^i}{\partial d_t^j} &= \gamma r \mathrm{Cov}_t^0\left[\bar{p}^i_{t+\Delta_t}, \bar{p}^j_{t+\Delta_t}\right] \\
&= \gamma r \mathrm{Cov}_t^0\Bigg[\frac{1}{2} f^i_{SS}\Delta S_1^2 + \Delta S_1\, O(\Delta_t) \\
&\qquad + O\big(\Delta_t^2\big), \frac{1}{2} f^j_{SS}\Delta S_2^2 + \Delta S_2\, O(\Delta_t) + O\big(\Delta_t^2\big)\Bigg] \\
&= \frac{\gamma r}{4} f^i_{SS} f^j_{SS} \mathrm{Cov}_t^0\left[\Delta S_1^2, \Delta S_2^2\right] + O\left(\Delta_t^{\frac{5}{2}}\right) \\
&= \frac{\gamma r}{2} f^i_{SS} f^j_{SS} \mathrm{Var}_t[\Delta S_1]\mathrm{Var}_t[\Delta S_2]\rho^2 + O\left(\Delta_t^{\frac{5}{2}}\right).
\end{aligned} \tag{40}
$$

Hence, since the option gammas f_{SS} are positive, we see that the cross-demand effect is positive regardless of the sign of the price correlation. The intuition for this surprising result is that the option dealer is hedged and, therefore, has profits or losses depending on the magnitude of the price changes, not their direction. Since the absolute price changes are positively correlated in this model, the unhedgeable risk is positively related, explaining the result.

For case (ii), we assume that assets have correlated stochastic volatilities and, for simplicity, that the stochastic volatilities are independent of the underlying prices. The price effect of demand is computed to be

$$\frac{\partial p_t^i}{\partial d_t^j} = \gamma r \mathrm{Cov}_t(\Delta\sigma_1, \Delta\sigma_2) f_\sigma^i f_\sigma^j + o(\Delta_t). \tag{41}$$

Making the reasonable assumption that $f_\sigma^i > 0$ and $f_\sigma^j > 0$, the sign of the price impact is the same as that of the correlation between σ_1 and σ_2. Hence, in this case, the demand for IBM options increases the price of Microsoft options if their volatilities are positively correlated and otherwise decreases the price of Microsoft options.

We could further extend the equilibrium determination of the number of dealers to the case in which they make markets in multiple underlyings (and possibly endogenize the number of underlyings dealers make markets in). Naturally, dealers enjoy the benefits of diversification and may additionally have economies of scale, at least up to a certain point. This would increase the equilibrium number of dealers, thus reducing the effect of demand pressure.

3. Empirical Results

The main focus of this article is the impact of net end-user option demand on option prices. We explore this impact empirically both for S&P 500 index options and for equity (i.e., individual stock) options.

3.1 Data

We acquire the data from three different sources. Data for computing net option demand were obtained directly from the Chicago Board Options Exchange (CBOE). These data consist of a daily record of closing short and long open interest on all SPX and equity options for public customers and firm proprietary traders from the beginning of 1996 to the end of 2001. We compute the net demand of each of these groups of agents as the long open interest minus the short open interest.

We focus our analysis on non-market-maker net demand defined as the sum of the net demand of public customers and proprietary traders, which is equal to the negative of the market-maker net demand (since options are in zero net supply). Hence, we assume that both public customers and firm proprietary traders—that is, all non-market-makers—are "end-users." We actually believe that proprietary traders are more similar to market-makers, and, indeed, their positions are more correlated with market-maker positions (the time-series correlation is 0.44). Consistent with this fact, our results are indeed stronger when we reclassify proprietary traders as market-makers (i.e., assume that end-users are the public customers). However, to be conservative and avoid any sample selection in favor of our predictions, we focus on the slightly weaker results. We note that, since proprietary traders constitute a relatively small group in our data, none of the main features of the descriptive statistics presented in this section or the results presented in the next section change under this alternative assumption.

The SPX options trade only at the CBOE while the equity options sometimes are cross-listed at other option markets. Our open interest data, however, include activity from all markets at which CBOE-listed options trade. The entire options market is comprised of public customers, firm proprietary traders, and market-makers so our data are comprehensive. For the equity options, we restrict attention to those underlying stocks with strictly positive option volume on at least 80% of the trade days over the 1996–2001 period. This restriction yields 303 underlying stocks.

The other main source of data for this article is the Ivy DB dataset from OptionMetrics LLC. The OptionMetrics data include end-of-day volatilities implied from option prices, and we use the volatilities implied from SPX and CBOE-listed equity options from the beginning of 1996 through the end of 2001. SPX options have European-style exercise, and OptionMetrics computes implied volatilities by inverting the Black-Scholes formula. When performing this inversion, the option price is set to the midpoint of the best closing bid and offer prices, the interest rate is interpolated from available LIBOR rates so that its maturity is equal to the expiration of the option, and the index dividend yield is determined from put-call parity. The equity options have American-style exercise, and OptionMetrics computes their implied volatilities using binomial trees that account for the early exercise feature and the timing and amount of the dividends expected to be paid by the underlying stock over the life of the options.

Finally, we obtain daily returns on the underlying index or stocks from the Center for Research in Security Prices (CRSP).

Definitions of variables: We refer to the difference between implied volatility and a reference volatility estimated from the underlying security as "excess implied volatility." This measures the option's "expensiveness," that is, its risk premium.

The reference volatility that we use for SPX options is the filtered volatility from the state-of-the-art model by Bates (2006), which accounts for jumps, stochastic volatility, and the risk premium implied by the equity market, but does not add extra risk premiums to (over-)fit option prices.[12] By subtracting the volatility from the Bates (2006) model, we account for the direct effects of jumps, stochastic volatility, and the risk premium implied by the equity market.

[12] We are grateful to David Bates for providing this measure.

Hence, excess implied volatility is the part of the option price unexplained by this model, which, according to our model, is due to demand pressure (and estimation error).

The reference volatility that we use for equity options is the predicted volatility over their lives from a GARCH(1,1) model estimated from five years of daily underlying stock returns leading up to the day of option observation.[13] (Alternative measures using historical or realized volatility lead to similar results.)

We conduct several tests on the time series ExcessImplVolATM of approximately ATM options with approximately one month to expiration. Specifically, for SPX, ExcessImplVolATM is the average excess implied volatility of options that have at least 25 contracts of trading volume, between 15 and 45 calendar days to expiration,[14] and moneyness between 0.99 and 1.01. (We compute the excess implied volatility variable only from reasonably liquid options in order to make it less noisy in light of the fact that it is computed using only one trade date.)

For equity options, ExcessImplVolATM is the average excess implied volatility of options with moneyness between 0.95 and 1.05, maturity between 15 and 45 calendar days, at least five contracts of trading volume, and implied volatilities available on OptionMetrics.

For the SPX, we also consider the excess implied-volatility skew ExcessImplVolSkew, defined as the implied-volatility skew over and above the skew predicted by the jumps and stochastic volatility of the underlying index. Specifically, the implied-volatility skew is defined as the average implied volatility of options with moneyness between 0.93 and 0.95 that trade at least 25 contracts on the trade date and have more than 15 and fewer than 45 calendar days to expiration, minus the average implied volatility of options with moneyness between 0.99 and 1.01 that meet the same volume and maturity criteria. In order to eliminate the skew that is due to jumps and stochastic volatility of the underlying, we consider the implied-volatility skew net of the similarly defined volatility skew implied by the objective distribution of Broadie, Chernov, and Johannes (2007), where the underlying volatility is that filtered from the Bates (2006) model.[15]

We consider four different demand variables for SPX options based on the aggregate net non-market-maker demand for options with 10–180 calendar days to expiration and moneyness between 0.8 and 1.20. First, NetDemand

[13] In particular, we use the GARCH(1,1) parameter estimates for the trade day (estimated on a rolling basis from the past five years of daily data) to compute the minimum mean square error volatility forecast for the number of trade days left in the life of the option. We annualize the volatility forecast (which is for the number of trade days left until the option matures) by multiplying by the square root of 252 and dividing by the square root of the number of trade days remaining in the life of the option.

[14] On any give trade day, these are the options with maturity closest to one month. Alternatively, for each month we could include in our test only the day that is precisely one month before expiration. This approach yields similar results.

[15] The model-implied skew is evaluated for one-month options with moneyness of, respectively, 0.94 and 1. We thank Mikhail Chernov for providing this time series.

is simply the sum of all net demands, which provides a simple atheoretical variable. The other three independent variables correspond to "weighting" the net demands using the models based on the market-maker risks associated with, respectively, discrete trading, jumps in the underlying, and stochastic volatility (Section 2.2). Specifically, DiscTrade weights the net demands by the Black-Scholes gamma as in the discrete-hedging model, JumpRisk weights by kappa computed using equally likely up and down moves of relative sizes 0.05 and 0.2 as in the jump model, and StochVol weights by maturity-adjusted Black-Scholes vega as in the stochastic volatility model. The Appendix provides more details on the computation of the model-based weighting factors. For equity options, we use just the NetDemand variable in the empirical work.

As a measure of skew in SPX-option demand, we use JumpRiskSkew, the excess implied-volatility skew from the demand model with underlying jumps described in Section 2.2. (We do not consider the models with discrete trading and stochastic volatility, since they do not have first-order skew implications, as explained in Section 2.2. We obtain similar, although weaker, results using an atheoretical measure based on raw demand.)

Furthermore, to test the robustness of our results, we also consider several control variables, which we motivate when we run the tests, in subsections 3.4 and 3.5. For the case of the SPX options, we consider three such variables. The first is the interaction between dealer profits (P&L) over the previous calendar month, calculated using dealer positions and assuming daily delta-hedging—as detailed in Section 3.3—and the measure of demand pressure. The other two variables are the current S&P 500 volatility filtered by Bates (2006) and the S&P 500 returns over the month leading up to the observation date.

For the case of equity options, the first control variable is the interaction between the number of option contracts traded on the underlying stock over the past six months, OptVolume and the measure of demand pressure. The other variables are the current volatility of the underlying stock measured from the past 60 trade days of underlying returns, the return on the underlying stock over the past month, as well as OptVolume on its own.

3.2 Descriptive statistics on end-user demand

Even though the SPX and individual equity option markets have been the subject of extensive empirical research, there is no systematic information on end-user demand in these markets. Panel A of Table 1 reports the average daily non-market-maker net demand for SPX options broken down by option maturity and moneyness (defined as the strike price divided by the underlying index level). Since our theoretical results indicate that the demand from a put or a call with the same strike price and maturity should have identical price impact, this table is constructed from the demands for puts and calls of all moneyness and maturity. For instance, the moneyness range 0.95–1 consists of put options that are up to 5% in-the-money and call options that are up to 5% OTM. Panel A indicates that 39% of the net demand comes from contracts

Table 1
Net demand for options by end-users

Mat. range (cal. days)	Moneyness range (K/S) 0–0.85	0.85–0.90	0.90–0.95	0.95–1.00	1.00–1.05	1.05–1.10	1.10–1.15	1.15–2.00	All
				Panel A: SPX option non-market-maker net demand					
1–9	6,014	1,780	1,841	2,357	2,255	1,638	524	367	16,776
10–29	7,953	1,300	1,115	6,427	2,883	2,055	946	676	23,356
30–59	5,792	745	2,679	7,296	1,619	−136	1,038	1,092	20,127
60–89	2,536	1,108	2,287	2,420	1,569	−56	118	464	10,447
90–179	7,011	2,813	2,689	2,083	201	1,015	4	2,406	18,223
180–364	2,630	3,096	2,335	−1,393	386	1,125	−117	437	8,501
365–999	583	942	1,673	1,340	1,074	816	560	−1,158	5,831
All	32,519	11,785	14,621	20,530	9,987	6,457	3,074	4,286	103,260
				Panel B: Equity option non-market-maker net demand					
1–9	−51	−25	−40	−45	−47	−31	−23	−34	−295
10–29	−64	−35	−57	−79	−102	−80	−55	−103	−576
30–59	−55	−31	−39	−55	−88	−90	−72	−144	−574
60–89	−47	−29	−37	−47	−60	−60	−55	−133	−469
90–179	−85	−60	−73	−84	−105	−111	−101	−321	−941
180–364	53	−19	−23	−24	−36	−35	−33	−109	−225
365–999	319	33	25	14	12	7	9	−56	363
All	70	−168	−244	−320	−426	−400	−331	−899	−2717

Average non-market-maker net demand for put and call option contracts for SPX and individual equity options by moneyness and maturity, 1996–2001. Equity-option demand is per underlying stock.

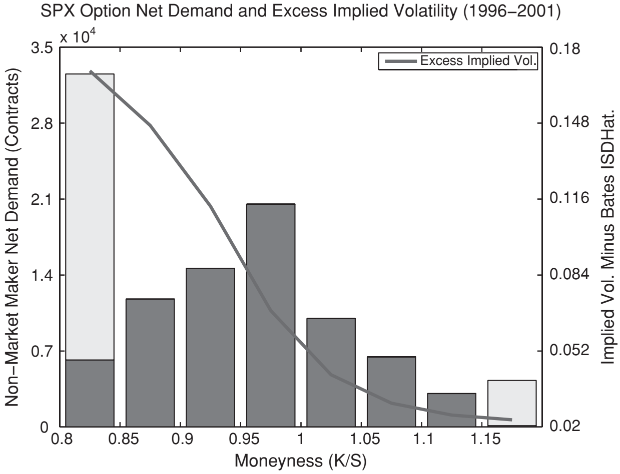

Figure 1
Index options: demand and expensiveness, measured in terms of implied volatility
The bars show the average daily net demand for puts and calls from non-market-makers for SPX options in the different moneyness categories (left axis). The top part of the leftmost (rightmost) bar shows the net demand for all options with moneyness less than 0.8 (greater than 1.2). The line is the average SPX excess implied volatility, that is, implied volatility minus the volatility from the underlying security filtered using Bates (2006), for each moneyness category (right axis). The data cover 1996–2001.

with fewer than 30 calendar days to expiration. Consistent with conventional wisdom, the good majority of this net demand is concentrated at moneyness where puts are OTM (i.e., moneyness <1.) Panel B of Table 1 reports the average option net demand per underlying stock for individual equity options from non-market-makers. With the exception of some long maturity option categories (i.e., those with more than one year to expiration and in one case with more than six months to expiration), the non-market-maker net demand for all of the moneyness/maturity categories is negative. That is, non-market-makers are net suppliers of options in all of these categories. This stands in a stark contrast to the index-option market in Panel A where non-market-makers are net demanders of options in almost every moneyness/maturity category.

Figure 1 illustrates the SPX option net demands across moneyness categories and compares these demands to the expensiveness of the corresponding options. The line in the figure plots the average SPX excess implied volatility for eight moneyness intervals over the 1996–2001 period. In particular, on each trade date the average excess implied volatility is computed for all puts and calls in a moneyness interval. The line depicts the means of these daily averages. The excess implied volatility inherits the familiar downward sloping smirk in SPX option-implied volatilities. The bars in Figure 1 represent the average daily net demand from non-market-maker for SPX options in the moneyness categories, where the top part of the leftmost (rightmost) bar shows the net demand for all options with moneyness less than 0.8 (greater than 1.2).

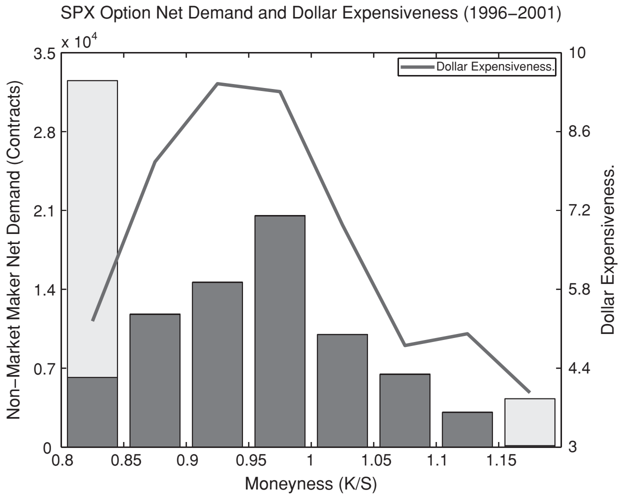

Figure 2
Index options: demand and expensiveness, measured in terms of dollars
The bars show the average daily net demand for puts and calls from non-market-makers for SPX options in the different moneyness categories (left axis). The top part of the leftmost (rightmost) bar shows the net demand for all options with moneyness less than 0.8 (greater than 1.2). The line is the average SPX price expensiveness, that is, the option price minus the "fair price" implied by Bates (2006) using filtered volatility from the underlying security and equity risk premiums, for each moneyness category (right axis). The data cover 1996–2001.

The first main feature of Figure 1 is that index options are expensive (i.e., have a large risk premium), consistent with what is found in the literature, and that end-users are net buyers of index options. This is consistent with our main hypothesis: end-users buy index options and market-makers require a premium to deliver them.

The second main feature of Figure 1 is that the net demand for low-strike options is greater than the demand for high-strike options. This could help explain the fact that low-strike options are more expensive than high-strike options (Proposition 4).

The shape of the demand across moneyness is clearly different from the shape of the expensiveness curve. This is expected for two reasons. First, our theory implies that demand pressure in one moneyness category impacts the implied volatility of options in other categories, thus "smoothing" the implied-volatility curve and changing its shape. Second, our theory implies that demands (weighted by the variance of the unhedgeable risks) affect prices, and the price effect must then be translated into volatility terms. It follows that a left-skewed hump-shaped price effect typically translates into a downward sloping volatility effect, consistent with the data. In fact, the observed average demands can give rise to a pattern of expensiveness similar to the one observed empirically when using a version of the model with jump risk. It is helpful to link these demands more directly to the predictions of our theory. Our model shows that every option contract demanded leads to an increase in its price—in dollar

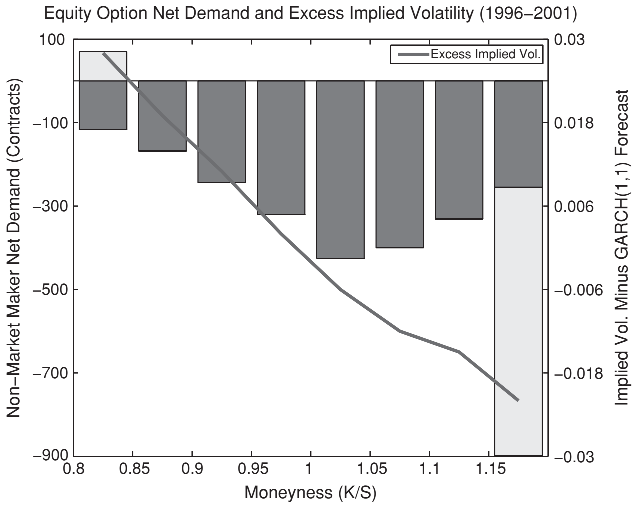

Figure 3
Equity options: demand and expensiveness, measured in terms of implied volatility
The bars show the average daily net demand per underlying stock from non-market-makers for equity options in the different moneyness categories (left axis). The top part of the leftmost (rightmost) bar shows the net demand for all options with moneyness less than 0.8 (greater than 1.2). The line is the average equity option excess implied volatility, that is, implied volatility minus the GARCH(1,1) expected volatility, for each moneyness category (right axis). The data cover 1996–2001.

terms—proportional to the variance of its unhedgeable part (and an increase in the price of any other option proportional to the covariance of the unhedgeable parts of the two options). Hence, the relationship between raw demands (that is, demands not weighted according to the model) and expensiveness is more directly visible when expensiveness is measured in dollar terms, rather than in terms of implied volatility. This fact is confirmed by Figure 2. Indeed, the price expensiveness has a similar shape to the demand pattern. Because of the cross-option effects and the absence of the weighting factor (the covariance terms), we do not expect the shapes to be identical.[16]

Figure 3 illustrates equity option net demands across moneyness categories and compares them to their expensiveness. The line in the figure plots the average equity option excess implied volatility (with respect to the GARCH(1,1) volatility forecast) per underlying stock for eight moneyness intervals over the 1996–2001 period. In particular, on each trade date for each underlying stock the average excess implied volatility is computed for all puts and calls in a moneyness interval. These excess implied volatilities are averaged across underlying stocks on each trade day for each moneyness interval. The line depicts the means of these daily averages. The excess implied volatility line is downward sloping but only varies by about 5% across the moneyness categories.

[16] Even if the model-implied relationship involving dollar expensiveness is more direct, we follow the literature and concentrate on expensiveness expressed in terms of implied volatility. This can be thought of as a normalization that eliminates the need for explicit controls for the price level of the underlying asset.

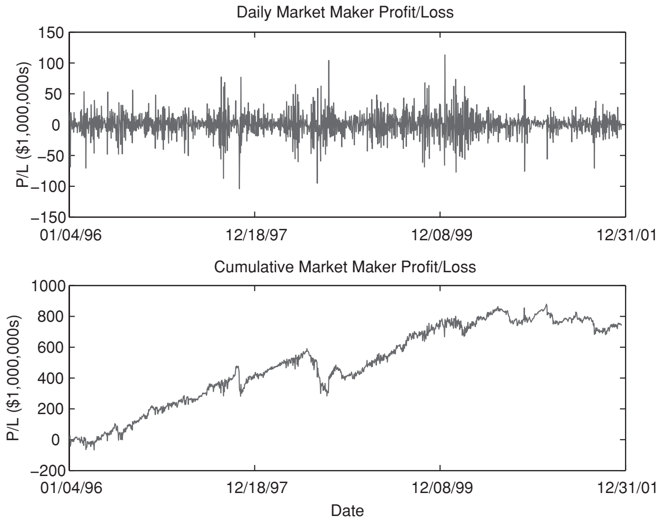

Figure 4
Option market-makers' estimated profits and losses from position taking
The top panel shows the market-makers' daily profits and losses (P&L) assuming they delta-hedge their option positions once per day. The bottom panel shows the corresponding cumulative P&Ls.

By contrast, for the SPX options the excess implied volatility line varies by 15% across the corresponding moneyness categories. The bars in the figure represent the average daily net demand per underlying stock from non-market-makers for equity options in the moneyness categories. The figure shows that non-market-makers are net sellers of equity options on average. Consistent with their different demand pattern, equity options do not appear expensive on average like index options.

3.3 Market-maker profits and losses

To illustrate the magnitude of the net demands, we compute approximate daily profits and losses (P&Ls) for the S&P 500 market-makers' hedged positions assuming daily delta-hedging. The daily and cumulative P&Ls are illustrated in Figure 4, which shows that the group of market-makers faces substantial risk that cannot be delta-hedged, with daily P&L varying between ca. \$100 million and \$−100 million. Further, the market-makers make cumulative profits of ca. \$800 million over the six-year period on their position taking.[17] With just over a hundred SPX market-makers on the CBOE, this corresponds to a profit of approximately \$1 million per year per market-maker.

Hence, consistent with the premise of our model, market-makers face substantial risk and are compensated on average for the risk that they take. Further,

[17] This number does not take into account the costs of market-making or the profits from the bid-ask spread on round-trip trades. A substantial part of market-makers' profit may come from the latter.

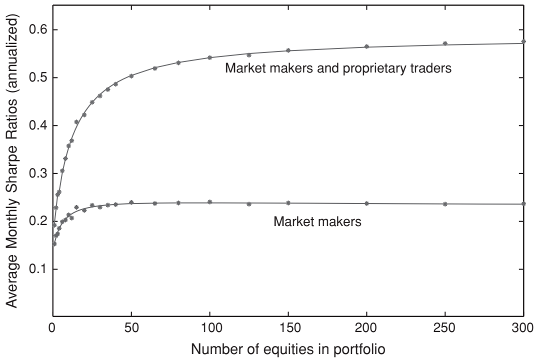

Figure 5
Option market-makers' estimated Sharpe ratio from position taking
The figure shows the annualized average monthly Sharpe ratios of the profits to making markets in options on a given number of individual equities (the horizontal axis). The profits are calculated using either the market-maker positions, or the combined market-maker and proprietary-trader positions.

consistent with the equilibrium entry of market-makers of Section 2.3, the market-makers' profits of about \$1 million per year per market-maker appear to be of the same order of magnitude as their cost of capital tied up in the trade, trader salaries, and back office expenses. In fact, the profit number appears low, but one must remember that market-makers likely make substantial profits from the bid-ask spread, an effect that we do not include in our profit calculations.

Another measure of the market-makers' compensation for accommodating SPX option demand pressures is the annualized Sharpe ratio of their profits or losses. This measure is 0.41 when computed from the daily P&L, an unimpressive risk/reward tradeoff comparable to that of a passive investment in the overall (stock) market. Since the daily P&L is negatively autocorrelated, the annualized Sharpe ratio increases to 0.85 when computed from monthly P&L (and hardly increases if we aggregate over longer time horizons). This Sharpe ratio reflects compensation for the risk that market-makers bear, as well as the committed capital that has alternative productive use and the dealers' effort and skill. This Sharpe ratio is of a magnitude consistent with equilibrium entry of market-makers as modeled in Section 2.3.

To illustrate the profits from making markets in individual-equity options, as well as the value from diversifying across a number of different underlying equities, we construct the annualized average monthly Sharpe ratios for hypothetical market-makers that deal options on n different stocks, with n varying from 1 to 300. Figure 5 shows the results when using both only the market-maker, and the combined market-maker and proprietary-trader positions as measures of the demand. The Sharpe ratios are, as expected, positive and increasing in the number of stocks in which a market-maker deals options. This provides further evidence that market-makers are compensated for providing liquidity, and that this liquidity provision is associated with a significant risk, part of which is diversifiable.

3.4 Net demand and expensiveness: SPX index options

Theorem 2 relates the demand for any option to a price impact on any option. Since our data contain both option demands and prices, we can test these theoretical results directly. Doing so requires that we choose a reason for the underlying asset and risk-free bond to form a dynamically incomplete market, and, hence, for the weighting factors $\mathrm{Cov}_t(\bar{p}^i_{t+1}, \bar{p}^j_{t+1})$ to be nonzero. As sources of market incompleteness, we consider discrete-time trading, jumps, and stochastic volatility using the results derived in Section 2.2.

We test the model's ability to help reconcile the two main puzzles in the option literature, namely the drivers of the overall level of implied volatility and its skew across option moneyness. The first set of tests investigates whether the overall excess implied volatility is higher on trade dates where the demand for options—aggregated according to the model—is higher. The second set of tests investigates whether the excess implied-volatility skew is steeper on trade dates where the model-implied demand-based skew is steeper.

Level: We investigate first the time-series evidence for Theorem 2 by regressing a measure of excess implied volatility on one of various demand-based explanatory variables:

$$\mathrm{ExcessImplVolATM}_t = a + b \text{ Demand Variable}_t + \varepsilon_t. \tag{42}$$

We run the regression on a monthly basis by averaging demand and expensiveness over each month. We do this to avoid day-of-the-month effects. (Our results are similar in an unreported daily regression.)

The results are shown in Table 2. We report the results over two subsamples because there are reasons to suspect a structural change in 1997. The change, apparent also in the time series of open interest and market-maker and public-customer positions (not shown here), stems from several events that altered the market for index options in the period from late 1996 to October 1997, such as the introduction of S&P 500 e-mini futures and futures options on the competing Chicago Mercantile Exchange (CME), the introduction of Dow Jones options on the CBOE, and changes in margin requirements. Our results are robust to the choice of these sample periods.[18]

We see that the estimate of the demand effect b is positive but insignificant over the first subsample, and positive and statistically significant

[18] We repeated the analysis with the sample periods lengthened or shortened by two or four months. This does not change our results qualitatively.

Table 2
Index-option expensiveness explained by end-user demand

	Before structural changes 1996/01–1996/10				After Structural changes 1997/10–2001/12			
Constant	0.0001 (0.004)	0.0065 (0.30)	0.005 (0.17)	0.020 (0.93)	0.04 (7.28)	0.033 (4.67)	0.032 (7.7)	0.038 (7.4)
NetDemand	2.1×10^{-7} (0.87)				3.8×10^{-7} (1.55)			
DiscTrade		6.9×10^{-10} (0.91)				2.8×10^{-9} (3.85)		
JumpRisk			6.4×10^{-6} (0.79)				3.2×10^{-5} (3.68)	
StochVol				8.7×10^{-7} (0.27)				1.1×10^{-5} (2.74)
Adj. R^2 (%)	9.6	6.1	8.1	0.5	7.0	18.5	25.9	15.7
N	10	10	10	10	50	50	50	50

The SPX excess implied volatility (i.e., observed implied volatility minus volatility from the Bates (2006) model) is regressed on the SPX non-market-maker demand pressure. The demand pressure is either (i) equal weighted demand across contracts (NetDemand), or demand weighted using our model in which market-maker risk is due to (ii) discrete-time trading risk (DiscTrade), (iii) jump risk (JumpRisk), or (iv) stochastic-volatility risk (StochVol). t-statistics computed using Newey-West are in parentheses.

over the second, longer, subsample for all three model-based explanatory variables.[19]

The expensiveness and the fitted values from the jump model are plotted in Figure 6, which clearly shows their comovement over the later sample. The fact that the b coefficient is positive indicates that, on average, when SPX net demand is higher (lower), SPX excess implied volatilities are also higher (lower). For the most successful model, the one based on jumps, changing the dependent variable from its lowest to its highest values over the late subsample would change the excess implied volatility by about 5.6 percentage points. A one-standard-deviation change in the jump-based demand variable results in a one-half-standard-deviation change in excess implied volatility (the corresponding R^2 is 26%). The model is also successful in explaining a significant proportion of the level of the excess implied volatility. Over the late subsample, demand explains on average 1.7 percentage points of excess implied volatility given the average demand and the regression coefficient, more than a third of the estimated average level of excess implied volatility.

We note that, in addition to the CBOE demand pressure observed in our data, there is over-the-counter demand for index options, for instance, via such products for individual investors as index-linked bonds. These securities give end-users essentially a risk-free security in combination with a call option on the index (or the index plus a put option), which leaves Wall Street short index options. Of course, this demand also contributes to the excess implied volatility.

[19] The model-based explanatory variables work better than just adding all contracts (NetDemand), because they give greater weight to near-the-money options. If we just count contracts using a more narrow band of moneyness, then the NetDemand variable also becomes significant.

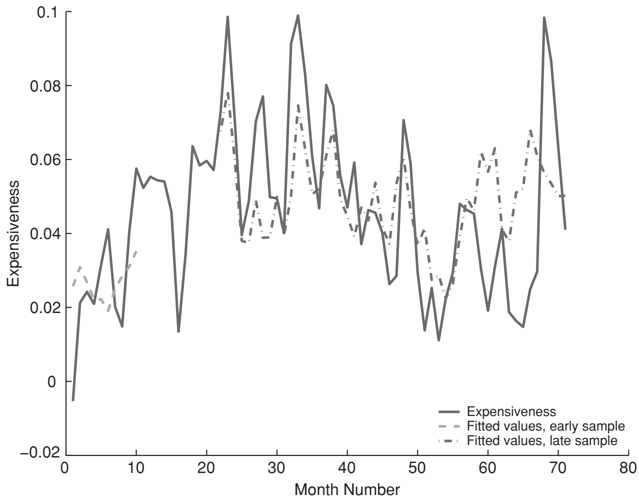

Figure 6
Actual and model-implied expensiveness of index options
The solid line shows the expensiveness of SPX options, that is, implied volatility of one-month at-the-money options minus the volatility measure of Bates (2006), which takes into account jumps, stochastic volatility, and the risk premium from the equity market. The dashed lines are, respectively, the fitted values of demand-based expensiveness using a model with underlying jumps, before and after certain structural changes (1996/01–1996/10 and 1997/10–2001/12).

Hence, our estimate that demand pressure explains on average 1.7 percentage points (and 26% of the variation) of the excess implied volatility can be viewed as a conservative lower bound.

Further support for the hypothesis that the supply for options is upward sloping comes from the comparison between the estimated supply-curve slopes following market-maker losses and, respectively, gains. If market-maker risk aversion plays an important role in pricing options, then one would expect prices to be less sensitive to demand when market-makers are well funded following profitable periods. This is exactly what we find. Breaking the daily sample[20] into two subsamples depending on whether the hedged market-maker profits over the previous 20 trading days is positive or negative,[21] we estimate the regression (42) for each subsample and find that, following losses, the b coefficient is approximately twice as large as the coefficient obtained in the other subsample. For instance, in the jump model, the regression coefficient following losses is 2.6E-05 with a t-statistic of 3.7 (330 observations), compared to a value of 1.1E-5 with a t-statistic of 6.1 in the complementary subsample (646 observations).

This effect of recent profits and losses (P&L) can also be studied in a time-series regression with the recent P&L on the right-hand side interacted with

[20] Because of the structural changes discussed above, we restrict our attention to the period starting on October 1, 1997.

[21] Similar results obtain if the breaking point is the mean or median daily profit.

Table 3
Index-option expensiveness explained by end-user demand with control variables

	Non-market-maker demand		
Constant	0.020	0.021	0.044
	(2.75)	(2.79)	(1.61)
Demand	4.23×10^{-5}	4.03×10^{-5}	4.16×10^{-5}
	(4.20)	(3.86)	(4.44)
P&L × Demand		-8.43×10^{-14}	-1.25×10^{-13}
		(−1.22)	(−1.73)
Volatility			-1.42×10^{-1}
			(−0.88)
S&P Return			8.65×10^{-3}
			(0.23)
Adj. R^2 (%)	31.0	31.9	34.3

The SPX excess implied volatility is regressed on non-market-maker demand as well as control variables, 1997/10–2001/12. The demand across contracts is weighted using our model with jump risk. The controls are (i) the product between lagged monthly market-maker profit and demand, (ii) current S&P 500 volatility, and (iii) the lagged monthly S&P 500 return. t-statistics computed using Newey-West are in parentheses. Demand has a positive effect on implied volatility, and the negative coefficient on the interaction between market-maker profits and demand pressure means that the effect of demand is larger following market-maker losses.

the magnitude of demand. Table 3 reports the result of including this and other control variables, namely the volatility of the underlying and the past return on the S&P. In order to save space, the only method of weighting demand that we present is the one given by the jump model and we restrict attention to the longer subsample.

First, the table makes it clear that, regardless of the additional controls used, demand pressure has a significant positive effect on the level of the implied-volatility curve. Further, there is a negative coefficient on the interaction between market-maker profits and demand pressure. This means that the effect of demand is larger following market-maker losses (i.e., negative profits) and smaller following positive profits. As explained above, this is consistent with the idea that following losses, market-makers are more risk averse and therefore the option prices are more sensitive to demand. Hence, the negative coefficient offers further support for our demand-based option pricing.

Skew: We next investigate the explanatory power of the model for the "skew" of the implied-volatility curve by regressing a measure of the steepness of the implied-volatility skew on our demand-based explanatory variable:

$$\text{ExcessImplVolSkew}_t = a + b\,\text{JumpRiskSkew}_t + \varepsilon_t. \tag{43}$$

Column 1 of Table 4 reports the monthly OLS estimates of this skewness regression. As expected, the model-implied effect of demand on the implied-volatility skew has a positive coefficient. To illustrate the magnitude of the effect, we note that a one-standard-deviation move in the independent variable results in a change in the dependent variable of 0.53 standard deviations. Further, Figure 7 illustrates graphically the demand effect on the volatility skew. (As discussed above, we divide the sample into two subsamples because

Table 4
Index option skew explained by end-user demand

	Non-market-maker demand		
Constant	0.032	0.033	0.005
	(13.23)	(11.10)	(0.50)
Demand	1.81×10^{-5}	1.55×10^{-5}	1.58×10^{-5}
	(2.99)	(2.73)	(3.40)
P&L × Demand		-1.12×10^{-13}	-6.67×10^{-14}
		(−2.96)	(−1.76)
Volatility			1.33×10^{-1}
			(2.80)
S&P Return			5.04×10^{-2}
			(2.96)
Adj. R^2 (%)	26.7	33.2	46.2

The actual SPX implied-volatility skew is regressed on demand-based skew implied by the non-market-maker demand and the jump-risk model, in the presence of control variables. The controls are (i) the product between lagged monthly market-maker profit and demand, (ii) current S&P 500 volatility, and (iii) the lagged monthly S&P 500 return. t-statistics computed using Newey-West are in parentheses.

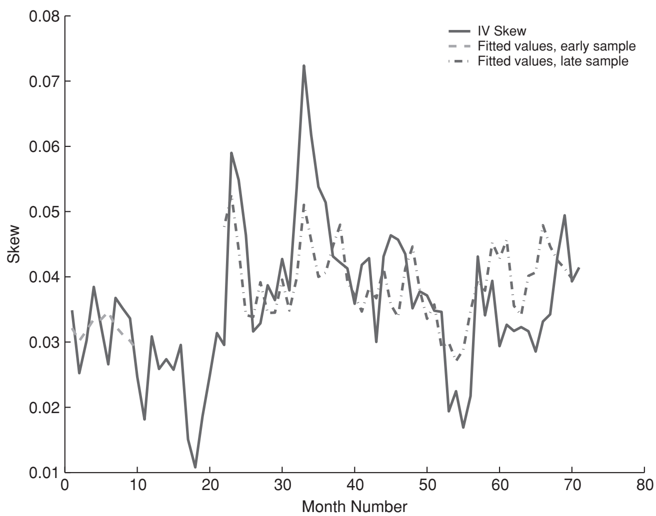

Figure 7
Actual and model-implied implied-volatility skew of index options
The solid line shows the implied-volatility skew for SPX options. The dashed lines are, respectively, the fitted values from the skew in demand before and after certain structural changes (1996/01–1996/10 and 1997/10–2001/12).

of structural changes and, while the figure shows both subsamples, the table focuses on the longer subsample to save space.)

Columns 2 and 3 of Table 4 include control variables. We see that the model-implied effect of demand on the implied-volatility skew has a positive coefficient in all specifications in confirmation of the model. Interestingly, the interaction of market-maker profits with the demand effect has a negative

Table 5
Equity option expensiveness explained by end-user demand

	Non-market-maker demand		
	1/96-12/01	1/96-6/99	10/99-12/01
Constant	0.078	0.072	0.085
	(13.58)	(12.34)	(7.84)
NetDemand	3.0×10^{-6}	4.1×10^{-6}	1.6×10^{-6}
	(7.66)	(11.64)	(4.62)
NetDemand × OptVolume	-1.0×10^{-12}	-1.8×10^{-12}	-7.9×10^{-14}
	(−3.91)	(−5.97)	(−1.34)
Volatility	−0.14	−0.16	−0.11
	(−14.23)	(−13.73)	(−10.51)
Return	−0.13	−0.12	−0.14
	(−11.57)	(−12.09)	(−6.16)
OptVolume	4.2×10^{-09}	5.9×10^{-09}	2.8×10^{-09}
	(1.98)	(2.11)	(0.89)
Adj. R^2 (%)	14.3	14.7	13.6

The relationship between excess-implied expensiveness—i.e., implied volatility minus GARCH volatility—and net demand for equity options on 303 different underlying stock (Equation 44), controlling for volatility, past return, option volume, and the interaction between net demand and option volume. We run monthly Fama-MacBeth cross-sectional regressions and report t-statistics adjusted for serial correlation via the Newey-West method.

coefficient. This is consistent with the idea that demand affects prices more after market-maker losses, when their risk aversion is likely to be higher.[22]

3.5 Net demand and expensiveness: equity options

We next investigate the equity option evidence for Theorem 2 by performing a cross-sectional Fama-MacBeth regression of a measure of excess implied volatility on the NetDemand explanatory variable:

$$\text{ExcessImplVolATM}_i = a + b\,\text{NetDemand}_i + \varepsilon_i. \tag{44}$$

We run the Fama-MacBeth regression on the 303 underlying stocks discussed above, namely those with strictly positive option volume on at least 80% of the trade days from the beginning of 1996 through the end of 2001. As in the case of the SPX options, we run the regressions on a monthly basis by averaging demand and expensiveness for each underlying stock for each month. (Once again, the results are similar for an unreported daily analysis.) Coefficient estimates and t-statistics (adjusted for serial correlation via Newey-West) are reported in Table 5. The average R^2s for the individual monthly cross-sectional regressions are also reported.

The first column of Table 5 shows the analysis for the entire sample period from the beginning of 1996 through the end of 2001. We see that the coefficient on the demand variable is positive and highly significant, controlling for volatility, past return, and option volume. In addition, the coefficient on the

[22] As a further robustness check, we confirm that the demand effect on the skew is also present when we use an atheoretical demand-skew measure based on aggregating the raw demands for low strikes, minus those with strikes close to 1. (This regression is not reported for brevity.)

interaction of demand with the level of option volume on the underlying stock is significantly negative. This result indicates that the demand effect is weaker when there is more option activity on a stock. This finding provides evidence in favor of our hypothesis, because greater option activity should be positively correlated with more capital being devoted to option market-making and, consequently, a smaller price impact per unit of option demand. This conclusion follows directly from our Proposition 5: part 1(a) shows that more expected demand leads to entry of more dealers, lowering the demand effects. Likewise, part 2 of the proposition shows that lower cost of being a dealer leads to more dealers, and additional volume leads to lower opportunity cost (denoted M in the proposition) because of additional profits to be made from earning the bid-ask spread on round-trip trades.

Columns 2 and 3 of Table 5 rerun the analysis separately on the subsamples before and after the summer of 1999. We perform the analysis on these subsamples because most options were listed only on one exchange before the summer of 1999, but many were listed on multiple exchanges after this period. Hence, there was potentially a larger total capacity for risk taking by market-makers after the cross listing. See, for instance, De Fontnouvelle, Fishe, and Harris (2003) for a detailed discussion of this well-known structural break. We follow De Fontnouvelle, Fishe, and Harris (2003) in the specific choice of the starting and ending dates for our subsamples, but we have confirmed that our results are robust to variations in this choice. For example, if the breakpoint for either subsample is moved by two months in either direction, there are only very small changes to the numbers reported in Table 5.

The main results are present in both the earlier and the later subperiod, although weaker in the later sample (indeed, the negative coefficient on the interaction term is no longer significant in the second part of the sample).[23] The demand coefficient is smaller in the second subsample when there was a greater capacity for risk taking by the market-makers due to cross-listing, another confirmation of the demand-based theory of option pricing. Finally, in unreported results we have run analogous Fama-MacBeth skew regressions on the equity options and also get significant results in the expected direction.

4. Conclusion

Relative to the Black-Scholes-Merton benchmark, index and equity options display a number of robust pricing anomalies. A large body of research has attempted to address these anomalies, in large part by generalizing the Black-Scholes-Merton assumptions about the dynamics of the underlying asset. While these efforts have met with undeniable success, nontrivial pricing puzzles remain. Further, it is not clear that this approach can yield a satisfactory description of option prices. For example, index and equity option prices display very

[23] When interpreting the t-statistics for the second subperiod, it should be borne in mind that it consists of only 27 monthly cross-sectional regressions.

different properties, even though the dynamics of their underlying assets are quite similar.

This article takes a different approach to option pricing. We recognize that, in contrast to the Black-Scholes-Merton framework, in the real world options cannot be hedged perfectly. Consequently, if intermediaries who take the other side of end-user option demand are risk averse, end-user demand for options will impact option prices.

The theoretical part of the article develops a model of competitive risk-averse intermediaries who cannot perfectly hedge their option positions. We compute equilibrium prices as a function of net end-user demand and show that demand for an option increases its price by an amount proportional to the variance of the unhedgeable part of the option and that it changes the prices of other options on the same underlying asset by an amount proportional to the covariance of their unhedgeable parts.

The empirical part of the article measures the expensiveness of an option as its Black-Scholes implied volatility minus a proxy for the expected volatility over the life of the option. We show that on average index options are quite expensive by this measure, and that they have high positive end-user demand. Equity options, on the other hand, do not appear expensive on average and have a small negative end-user demand. In accordance with the predictions of our theory, we find that index options are overall more expensive when there is more end-user demand and that the expensiveness skew across moneyness is positively related to the skew in end-user demand across moneyness. In addition, demand effects are stronger for index options after recent market-maker losses than after market-maker gains. For equity options, we find a positive cross-sectional relationship between option expensiveness and end-user demand. In addition, the relationship is stronger when there is less option activity—and therefore presumably less option market-maker capacity for risk-taking—on an underlying stock.

Appendix

Appendix: Proofs

Proof of Lemma 1. We start by imposing the following technical conditions. First, d and X have compact supports. Second, the transition function π of X has the following regularity property.

Assumption 1. *Whenever $x^n \to x$, $\pi(\cdot, x^n) \to \pi(\cdot, x)$ uniformly under the total-variation norm.*

Finally, we require that R^e_{t+1} be bounded, to so that all expectations are well defined. The Bellman equation is

$$J(W_t; t, X_t) = -\frac{1}{k}e^{-k(W_t + G_t(d_t, X_t))}$$
$$= \max_{C_t, q_t, \theta_t} \left\{ -\frac{1}{\gamma}e^{-\gamma C_t} + \rho \mathrm{E}_t \left[J(W_{t+1}; t+1, X_{t+1}) \right] \right\}. \quad \text{(A1)}$$

Given the strict concavity of the utility function, the maximum is characterized by the first-order conditions (FOCs). Using the proposed functional form for the value function, the FOC for C_t is

$$0 = e^{-\gamma C_t} + kR_f \rho \mathrm{E}_t \left[J(W_{t+1}; t+1, X_{t+1}) \right], \tag{A2}$$

which, together with (A1), yields

$$0 = e^{-\gamma C_t} + kR_f \left[J(W_t; t, X_t) + \frac{1}{\gamma} e^{-\gamma C_t} \right], \tag{A3}$$

that is,

$$e^{-\gamma C_t} = e^{-k(W_t + G_t(d_t, X_t))}, \tag{A4}$$

implying (6). The FOCs for θ_t and q_t are (7) and (8). We derive G recursively as follows. First, we let $G(t+1, \cdot)$ be given. Then, θ_t and q_t are given as the unique solutions to Equations (7) and (8). Clearly, θ_t and q_t do not depend on the wealth W_t. Further, (A3) implies that

$$0 = e^{-\gamma C_t} - R_f \rho \mathrm{E}_t \left[e^{-k\left((W_t - C_t)R_f + q_t(p_{t+1} - R_f p_t) + \theta_t R^e_{t+1} + G_{t+1}(d_{t+1}, X_{t+1})\right)} \right], \tag{A5}$$

that is,

$$e^{-\gamma C_t - kR_f C_t + kR_f W_t} = R_f \rho \mathrm{E}_t \left[e^{-k\left(q_t(p_{t+1} - R_f p_t) + \theta_t R^e_{t+1} + G_{t+1}(d_{t+1}, X_{t+1})\right)} \right], \tag{A6}$$

which, using (6), yields the equation that defines $G_t(d_t, X_t)$ (since X_t is Markov):

$$e^{-kR_f G_t(d_t, X_t)} = R_f \rho \mathrm{E}_t \left[e^{-k\left(q_t(p_{t+1} - R_f p_t) + \theta_t R^e_{t+1} + G_{t+1}(d_{t+1}, X_{t+1})\right)} \right]. \tag{A7}$$

In the online appendix we further prove the existence of a stationary solution at $t = \overline{T}$. ■

Proof of Proposition 1. Given a position process from date t onward and a price process from date $t+1$ onward, the price at time t is determined by (8). It is immediate that p_t is measurable with respect to time-t information. ■

Proof of Proposition 2. (i) is immediate, since prices are linear. Part (ii) follows because, for any $a \in \mathbb{R}$, the pricing kernel is kept exactly the same by the offsetting change in (q, θ). ■

Proof of Theorem 2. We start by calculating explicitly the sensitivity of the prices of a derivative p^i_t with respect to the demand pressure of another derivative d^j_t. We can initially differentiate with respect to q rather than d since $q^i = -d^i_t$.

For this, we first differentiate the pricing kernel:[24]

$$\frac{\partial m^d_{t+1}}{\partial q^j_t} = -km^d_{t+1} \left(p^j_{t+1} - R_f p^j_t + \frac{\partial \theta_t}{\partial q^j_t} R^e_{t+1} \right), \tag{A8}$$

using the facts that $\frac{\partial G(t+1, X_{t+1}; q)}{\partial q^j_t} = 0$ and $\frac{\partial p_{t+1}}{\partial q^j_t} = 0$. With this result, it is straightforward to differentiate (16) to get

$$0 = \mathrm{E}_t \left(m^d_{t+1} \left(p^j_{t+1} - R_f p^j_t + \frac{\partial \theta_t}{\partial q^j_t} R^e_{t+1} \right) R^e_{t+1} \right), \tag{A9}$$

[24] We suppress the arguments of functions. We note that p_t, θ_t, and G_t are functions of (d_t, X_t, t), and m^d_{t+1} is a function of $(d_t, X_t, d_{t+1}, X_{t+1}, y_{t+1}, R^e_{t+1}, t)$.

which implies that the marginal hedge position is

$$\frac{\partial \theta_t}{\partial q_t^j} = -\frac{\mathrm{E}_t\left(m_{t+1}^d\left(p_{t+1}^j - R_f p_t^j\right) R_{t+1}^e\right)}{\mathrm{E}_t\left(m_{t+1}^d\left(R_{t+1}^e\right)^2\right)} = -\frac{\mathrm{Cov}_t^d\left(p_{t+1}^j, R_{t+1}^e\right)}{\mathrm{Var}_t^d\left(R_{t+1}^e\right)}. \tag{A10}$$

Similarly, we derive the price sensitivity by differentiating (15):

$$\begin{aligned}
\frac{\partial p_t^i}{\partial q_t^j} &= -k\mathrm{E}_t\left[m_{t+1}^d\left(p_{t+1}^j - R_f p_t^j + \frac{\partial \theta_t}{\partial q_t^j} R_{t+1}^e\right) p_{t+1}^i\right] \\
&= -\frac{k}{R_f}\mathrm{E}_t^d\left[\left(p_{t+1}^j - R_f p_t^j - \frac{\mathrm{Cov}_t^d\left(p_{t+1}^j, R_{t+1}^e\right)}{\mathrm{Var}_t^d\left(R_{t+1}^e\right)} R_{t+1}^e\right) p_{t+1}^i\right] \\
&= -\gamma(R_f - 1)\mathrm{E}_t^d\left[\bar{p}_{t+1}^j \bar{p}_{t+1}^i\right] \\
&= -\gamma(R_f - 1)\mathrm{Cov}_t^d\left[\bar{p}_{t+1}^j, \bar{p}_{t+1}^i\right],
\end{aligned} \tag{A11}$$

where $\bar{p}_{t+1}^i$ and $\bar{p}_{t+1}^j$ are the unhedgeable parts of the price changes defined in the text. ■

Proof of Proposition 3. Part (a) is immediate since a variance is always positive. The proof of (b) is based on the following result, which is proved in the online appendix.

Lemma 2. *Given h_1 and h_2 convex functions on $\mathbb{R}$, $\forall \beta < 0, \alpha, \gamma \in \mathbb{R}$, $\exists\, \alpha', \gamma' \in \mathbb{R}$ such that*

$$|h_1(x) - \alpha' x - \gamma'| \le |h_1(x) - \alpha x - \beta h_2(x) - \gamma| \tag{A12}$$

$\forall x \in \mathbb{R}$. Consequently, under any distribution, regressing h_1 on h_2 and the identity function results in a positive coefficient on h_2.

Letting $\tilde{p}_{t+1} = p_{t+1} - \mathrm{E}_t^d\left[p_{t+1}\right]$ and suppressing subscripts, consider the expression

$$\Psi = \mathrm{E}^d\left[\tilde{p}^i \tilde{p}^j\right]\mathrm{Var}(R^e) - \mathrm{E}^d\left[\tilde{p}^i R^e\right]\mathrm{E}^d\left[\tilde{p}^j R^e\right], \tag{A13}$$

which we want to show to be positive. Letting $\hat{p}^i = \mathrm{E}^d[\tilde{p}^i \mid S]$ and $\hat{p}^j = \mathrm{E}^d[\tilde{p}^j \mid S]$, we write

$$\begin{aligned}
\Psi &= \mathrm{E}^d\left[\mathrm{Cov}\left(\tilde{p}^i, \tilde{p}^j \mid S\right)\mathrm{Var}(R^e) + \hat{p}^i \hat{p}^j \mathrm{Var}(R^e) - \mathrm{E}^d\left[\hat{p}^i R^e\right]\mathrm{E}^d\left[\hat{p}^j R^e\right]\right] \\
&= \mathrm{E}^d\left[\mathrm{Cov}\left(\tilde{p}^i, \tilde{p}^j \mid S\right)\mathrm{Var}(R^e)\right] + \mathrm{E}^d\left[\hat{p}^i \hat{p}^j \mathrm{Var}(R^e) - \mathrm{E}^d\left[\hat{p}^i R^e\right]\mathrm{E}^d\left[\hat{p}^j R^e\right]\right].
\end{aligned} \tag{A14}$$

The first term is positive by assumption, while the second is positive because $\hat{p}^i$ and $\hat{p}^j$ are convex and therefore Lemma 2 applies. ■

Proof of Theorem 3. We compute the sensitivity of current prices to a deviation in future positions from 0 in the direction of demand $\tilde{d}_s = \epsilon_s d_s$ at time s by differentiating with respect to $\epsilon_s = \epsilon$ (evaluated at $\epsilon = 0$). We then aggregate the demands at all times to compute the total effect:

$$\frac{\partial p_t}{\partial \epsilon} = \sum_{s \ge 0} \frac{\partial p_t}{\partial \epsilon_s}\frac{\partial \epsilon_s}{\partial \epsilon} = \sum_{s \ge 0} \frac{\partial p_t}{\partial \epsilon_s}. \tag{A15}$$

To compute the price effect of expected demand at any time s, we note that it follows from the dealer's problem that

$$p_t = \mathrm{E}_t\left[\rho^{s-t} e^{-\gamma(C_s - C_t)} p_s\right], \tag{A16}$$

which implies

$$\frac{\partial p_t}{\partial \epsilon_s} = \mathrm{E}_t\left[\rho^{s-t} e^{-\gamma(C_s - C_t)} \frac{\partial p_s}{\partial \epsilon_s}\right] = R_f^{-(s-t)} \mathrm{E}_t^0\left[\frac{\partial p_s}{\partial \epsilon_s}\right] = R_f^{-(s-t)} \mathrm{E}_t^0\left[\frac{\partial p_s}{\partial \tilde{q}_s^j} q_s^j\right], \quad \text{(A17)}$$

where we use that $\frac{\partial C_t}{\partial \epsilon_s} = \frac{\partial C_s}{\partial \epsilon_s} = 0$ at $q = 0$. The equality $\frac{\partial C_s}{\partial \epsilon_s} = 0$ follows from

$$\frac{\partial C_s}{\partial q_s^j} = \frac{k}{\gamma} \frac{\partial G(s, X_s; q)}{\partial q_s^j} = -\frac{k^2 R_f \rho}{\gamma} \mathrm{E}_s^0\left[p_{s+1}^j - R_f p_s^j + \frac{\partial \theta_s}{\partial q_s^j} R_{s+1}^e\right] = 0 \quad \text{(A18)}$$

and the other equality follows from differentiating the condition that marginal rates of substitution are equal:

$$e^{-\gamma C_t} = e^{-\rho(s-t)} \mathrm{E}_t\left[e^{-\gamma C_s}\right], \quad \text{(A19)}$$

which gives

$$e^{-\gamma C_t} \frac{\partial C_t}{\partial \epsilon_s} = e^{-\rho(s-t)} \mathrm{E}_t\left[e^{-\gamma C_s} \frac{\partial C_s}{\partial \epsilon_s}\right] = 0. \quad \text{(A20)}$$

Finally, in the online appendix we show that the price is a smooth (C^∞) function of ϵ.

Proof of Proposition 4. Consider an optimally hedged short put position with strike price $K < R_f S_t$. With $x = S_{t+1} - R_f S_t$, the payoff from this position is

$$\Pi(x) = -d(K - R_f S_t - x)^+ + \theta x. \quad \text{(A21)}$$

The optimality of the hedge means that, under the risk-neutral measure,

$$\mathrm{E}\left[e^{-k\Pi(x)} x\right] = 0. \quad \text{(A22)}$$

Note that, since $K < R_f S_t$, $\Pi(x) < 0$ for $x > 0$ and $\Pi(x) > 0$ for $K - R_f S_t < x < 0$. Consequently, given the symmetry of x around 0 and the zero-expectation condition above, with ξ denoting the density of x,

$$\int_{K - R_f S_t}^{\infty} \left(e^{-k\Pi(x)} x - e^{-k\Pi(-x)} x\right) \xi(x)\, dx = -\int_0^{K - R_f S_t} \left(e^{-k\Pi(x)} x - e^{-k\Pi(-x)} x\right) \xi(x)\, dx < 0. \quad \text{(A23)}$$

It immediately follows that it cannot be true that $\Pi(-x) \geq \Pi(x)$ for all $x > |K - R_f S_t|$. In other words, for some value $x > |K - R_f S_t|$, $\Pi(-x) < \Pi(x)$, which then gives $d + \theta > -\theta$, or $|\theta| < \frac{1}{2}|d|$: the payoff is more sensitive to large downward movements in the underlying than to large upward movements. Thus, there exists $\bar{K}$ such that, for all $S_{t+1} < \bar{K}$,

$$\Pi(S_{t+1} - R_f S_t) < \Pi(-(S_{t+1} - R_f S_t)), \quad \text{(A24)}$$

implying that, whenever $K' < \bar{K}$ and $K'' = 2R_f S_t - K'$,

$$p(p, K', d) > p(c, K'', d) \quad \text{(A25)}$$

$$p(p, K', 0) = p(c, K'', 0), \quad \text{(A26)}$$

the second relation being the result of symmetry. ■

Proof of Proposition 5. Fix a demand process $\bar{d}$ and write G as $G(\gamma', \epsilon\bar{d}, \gamma, X)$. It is then easily seen that $G(\gamma', \epsilon\bar{d}, \gamma, X) = G(\gamma', \bar{d}, \epsilon\gamma, X)$. It is equally clear that, for a given distribution of asset prices, the vector $\gamma'(q(\gamma'), \theta(\gamma'))$ is independent of the value of γ'. This readily implies that, if the set of dealers I are in the market, then $\gamma^{-1} = \int_I \gamma'(i)^{-1} d\mu(i)$. It is also easily checked that the value $\gamma' G(\gamma', \epsilon\bar{d}, \gamma, X)$ is independent of γ', so that

$$G(\gamma', \epsilon\bar{d}, \gamma, X) = \frac{1}{\gamma'} G(1, \epsilon\bar{d}, \gamma, X).$$

We now show that G increases in ϵ. To that end we write the value function as

$$J_0 = -\mathrm{E}_0\left[\frac{1}{\gamma'}\sum_{t=0}^{T}\rho^t e^{-\gamma' C_t} + \frac{1}{k'}\rho^{T+1} e^{-k'(W_{T+1}+G'_{T+1})}\right], \tag{A27}$$

where

$$W_{T+1} = (W_0 - C_0)R_f^{T+1} + \sum_{t=0}^{T}\big(-C_{t+1} + \theta_t R^e_{t+1} - q_t(p_{t+1} - R_f p_t)\big)R_f^{T-t}$$

and $k' = \gamma' r/(r-1)$, and proceed to calculate the derivative $\frac{\partial J_0}{\partial \epsilon}$. From the FOCs,

$$\frac{\partial J_0}{\partial \epsilon} = -\mathrm{E}_0^d\left[\rho^{T+1}\sum_{t=0}^{T} q_t\left(\frac{\partial p_{t+1}}{\partial \epsilon} - R_f\frac{\partial p_t}{\partial \epsilon}\right)R_f^{T-t}\right],$$

with

$$\frac{\partial p_t}{\partial \epsilon} = E_t^d\left[\sum_{s=t}^{T} R_f^{-(s-t)}\bar{p}_{s+1}\bar{p}'_{s+1}d_s\right].$$

It follows that

$$\begin{aligned}\frac{\partial J_0}{\partial \epsilon} &\propto -\mathrm{E}_0^d\left[\sum_{t=0}^{T} q_t\left(\mathrm{E}_t^d\left[\sum_{s=t+1}^{T} R_f^{-(s-t-1)}\bar{p}_{s+1}\bar{p}'_{s+1}d_s - R_f\sum_{s=t}^{T}R_f^{-(s-t)}\bar{p}_{s+1}\bar{p}'_{s+1}d_s\right]\right)R_f^{T-t}\right]\\ &= -\mathrm{E}_0^d\left[\sum_{t=0}^{T} q_t\mathrm{E}_t^d\left[\bar{p}_{t+1}\bar{p}'_{t+1}\right]d_t R_f^{T+1-t}\right]\\ &= \mathrm{E}_0^d\left[\sum_{t=0}^{T} d_t\mathrm{E}_t^d\left[\bar{p}_{t+1}\bar{p}'_{t+1}\right]d_t R_f^{T+1-t}\right] > 0.\end{aligned}$$

Consider now dealer entry. A given dealer with risk aversion γ' enters the business if and only if

$$-\frac{1}{k'}e^{-k'(W_0 - M + G(\gamma', d, \gamma, X))} \geq -\frac{1}{k'}e^{-k'(W_0 + \bar{G}'(\gamma'))},$$

or

$$G(\gamma', d, \gamma, X) - \bar{G}'(\gamma') \geq M,$$

where from being able to invest in the underlying, but having no access to options. (This is the same as G' when demand is identically zero.) Using the results above, the entry condition becomes

$$G(1, \bar{d}, \epsilon\gamma, X) - \bar{G}'(1) \geq \gamma' M. \tag{A28}$$

It is immediate that, in any equilibrium, a dealer with risk aversion $\gamma'' > \gamma'$ is in the market if and only if all dealers with risk aversion γ' are. We may therefore assume that any equilibrium is characterized by dealers in $I = [0, \bar{i}]$. Letting now $\bar{i}$ vary from 0 to ∞, the resulting aggregate risk aversion $\gamma(\bar{i})$ decreases from infinity to 0. As $G(1, \epsilon\bar{d}, \gamma(\bar{i}), X) - \bar{G}'(1)$ decreases with $\bar{i}$ and tends to 0 as $\gamma(\bar{i})$ tends to 0, while $\gamma'(\bar{i})$ increases, a unique dealer-entry equilibrium exists.

For part 1 (a), we use the fact that $G(\gamma'(\bar{i}), \bar{d}, \epsilon\gamma(\bar{i}), X)$ increases in ϵ, and therefore so does the equilibrium $\bar{i}$. Furthermore, if γ' increases from one equilibrium value $\bar{i}$ to another one, then the right-hand side of (A28) increases, which means that so must the left-hand side, i.e., the product $\epsilon\gamma(\bar{i})$ must increase. Using Theorem 3,

$$\frac{dp}{d\epsilon} = \frac{d(\gamma(\epsilon)\epsilon)}{d\epsilon} v,$$

at $\epsilon = 0$, for some vector v. Since $\gamma(\epsilon)\epsilon$ increases with ϵ, the effect of increasing demand is to amplify price deviations from zero-demand levels, at least for ϵ close to 0. If $\gamma(\epsilon)\epsilon$ is actually constant—because $\gamma'(\bar{i})$ is—then the two equilibria are identical from a pricing perspective. This finishes the proof of part 1.

Part 2 follows immediately from the preceding analysis. ■

References

Amin, K., J. D. Coval, and H. N. Seyhun. 2004. Index Option Prices and Stock Market Momentum. *Journal of Business* 77:835–74.

Back, K. 1993. Asymmetric Information and Options. *Review of Financial Studies* 6:435–72.

Bakshi, G., N. Kapadia, and D. Madan. 2003. Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options. *Review of Financial Studies* 16:101–43.

Bates, D. S. 2000. Post 87 Crash Fears in the S&P 500 Futures Options Market. *Journal of Econometrics* 94:181–238.

Bates, D. S. 2003. Empirical Option Pricing: A Retrospection. *Journal of Econometrics* 116:387–404.

Bates, D. S. 2006. Maximum Likelihood Estimation of Latent Affine Processes. *Review of Financial Studies* 19:909–65.

Benzoni, L., P. Collin-Dufresne, and R. S. Goldstein. 2005. Can Standard Preferences Explain the Prices of Out of the Money S&P 500 Put Options? Working Paper, NBER.

Bernardo, A. E., and O. Ledoit. 2000. Gain, Loss and Asset Pricing. *The Journal of Political Economy* 108:144–72.

Black, F., and M. S. Scholes. 1973. The Pricing of Options and Corporate Liabilities. *The Journal of Political Economy* 81:637–54.

Bollen, N. P., and R. E. Whaley. 2004. Does Net Buying Pressure Affect the Shape of Implied Volatility Functions? *Journal of Finance* 59:711–53.

Bondarenko, O. 2003. Why Are Put Options So Expensive? Working Paper, University of Illinois at Chicago.

Brennan, M. 1979. The Pricing of Contingent Claims in Discrete Time Models. *Journal of Finance* 34:53–68.

Broadie, M., M. Chernov, and M. Johannes. 2007. Model Specification and Risk Premia: Evidence from Futures Options. *Journal of Finance* 62:1453–90.

Cochrane, J. H., and J. Saa-Requejo. 2000. Beyond Arbitrage: Good-Deal Asset Price Bounds in Incomplete Markets. *The Journal of Political Economy* 108:79–119.

Constantinides, G. M., J. C. Jackwerth, and S. Perrakis. 2009. Mispricing of S&P 500 Index Options. *Review of Financial Studies* 22:1247–277.

Constantinides, G. M., and S. Perrakis. 2002. Stochastic Dominance Bounds on Derivatives Prices in a Multiperiod Economy with Proportional Transaction Costs. *Journal of Economic Dynamics and Control* 26:1323–52.

Coval, J. D., and T. Shumway. 2001. Expected Option Returns. *Journal of Finance* 56:983–1009.

De Fontnouvelle, P., R. P. H. Fishe, and J. H. Harris. 2003. The Behavior of Bid-Ask Spreads and Volume in Options Markets during the Competition for Listings in 1999. *Journal of Finance* 58:2437–63.

de Roon, F., T. E. Nijman, and C. Veld. 2000. Hedging Pressure Effects in Futures Markets. *Journal of Finance* 55:1437–56.

Driessen, J., and P. Maenhout. 2007. An Empirical Portfolio Perspective on Option Pricing Anomalies. *Review of Finance* 11:561–603.

Easley, D., M. O'Hara, and P. S. Srinivas. 1998. Option Volume and Stock Prices: Evidence on Where Informed Traders Trade. *Journal of Finance* 53:432–65.

Figlewski, S. 1989. Options Arbitrage in Imperfect Markets. *Journal of Finance* 44:1289–311.

Gabaix, X., A. Krishnamurthy, and O. Vigneron. 2007. Limits of Arbitrage: Theory and Evidence from the Mortgage-Backed Securities Market. *Journal of Finance* 62:557–95.

Greenwood, R. 2005. Short-and Long-term Demand Curves for Stocks: Theory and Evidence on the Dynamics of Arbitrage. *Journal of Financial Economics* 75:607–49.

Hakansson, N. 1979. The Fantastic World of Finance: Progress and the Free Lunch. *The Journal of Financial and Quantitative Analysis* 14:717–34.

Hicks, J. R. 1939. *Value and Capital*. Cambridge: Oxford University Press.

Hugonnier, J., D. Kramkov, and W. Schachermayer. 2005. On Utility Based Pricing of Contingent Claims in Incomplete Markets. *Mathematical Finance* 15:203–12.

Jackwerth, J. 2000. Recovering Risk Aversion from Option Prices and Realized Returns. *Review of Financial Studies* 13:433–51.

Keynes, J. M. 1923. Some Aspects of Commodity Markets. *Manchester Guardian Commercial, European Reconstruction Series* Section 13:784–86.

Longstaff, F. A. 1995. Option Pricing and the Martingale Restriction. *Review of Financial Studies* 8:1091–124.

Merton, R. 1973. The Theory of Rational Option Pricing. *Bell Journal of Economics and Management Science* 4:141–83.

Newman, Y., and M. Rierson. 2004. Illiquidity Spillovers: Theory and Evidence from European Telecom Bond Issuance. Working Paper, Stanford Graduate School of Business.

Poteshman, A. M. 2001. Underreaction, Overreaction, and Increasing Misreaction to Information in the Options Market. *Journal of Finance* 56:851–76.

Rubinstein, M. 1976. The Valuation of Uncertain Income Streams and Pricing of Options. *Bell Journal of Economics* 7:407–25.

Rubinstein, M. 1994. Implied Binomial Trees. *Journal of Finance* 49:771–818.

Shleifer, A. 1986. Do Demand Curves for Stocks Slope Down? *Journal of Finance* 41:579–90.

Soner, H. M., S. E. Shreve, and J. Cvitanic. 1995. There Is No Nontrivial Hedging Portfolio for Option Pricing with Transaction Costs. *The Annals of Applied Probability* 5:327–55.

Stapleton, R. C., and M. G. Subrahmanyam. 1984. The Valuation of Multivariate Contingent Claims in Discrete Time Models. *Journal of Finance* 39:207–28.

Stein, J. 1989. Overreactions in the Options Market. *Journal of Finance* 44:1011–23.

Wurgler, J., and E. V. Zhuravskaya. 2002. Does Arbitrage Flatten Demand Curves for Stocks? *Journal of Business* 75:583–608.