

Are Options on Index Futures Profitable for Risk-Averse Investors? Empirical Evidence

GEORGE M. CONSTANTINIDES, MICHAL CZERWONKO,
JENS CARSTEN JACKWERTH, and STYLIANOS PERRAKIS*

ABSTRACT

American options on the S&P 500 index futures that violate the stochastic dominance bounds of Constantinides and Perrakis (2009) from 1983 to 2006 are identified as potentially profitable trades. Call bid prices more frequently violate their upper bound than put bid prices do, while violations of the lower bounds by ask prices are infrequent. In out-of-sample tests of stochastic dominance, the writing of options that violate the upper bound increases the expected utility of any risk-averse investor holding the market and cash, net of transaction costs and bid-ask spreads. The results are economically significant and robust.

WE IDENTIFY AMERICAN call and put options on the S&P 500 index futures from 1983 to 2006 that violate the stochastic dominance upper bounds of Constantinides and Perrakis (2007) as potentially profitable investment opportunities—“good sell” options. We then consider the utility enhancement that obtains from exploiting such violations by adopting the appropriate trading policy for a generic investor who holds only the market index and the risk-free asset. In the identification of both the violations and the trading policy, we recognize the potential early exercise of these American options. We allow for realistic trading conditions by using only observable information and by incorporating transaction costs, bid-ask spreads, and trading delays (by waiting one quote before entering the position).

The main contribution of our paper is to show that trading policies that exploit these violations lead to out-of-sample portfolio returns that stochastically dominate (in the second order) portfolio returns that do not exploit them. This

*Constantinides is at the University of Chicago and the National Bureau of Economic Research. Czerwonko is at Concordia University and McGill University. Jackwerth is at the University of Konstanz. Perrakis is at Concordia University. We thank Oleg Bondarenko; Wolfgang Buehler; Jim Hodder; Robert Merton; Ingmar Nolte; Myron Scholes; Sorin Sorescu; Suresh Sundaresan; Michael Wolf; Campbell Harvey (the Editor); the anonymous associate editor and referees; participants at the Second Mont Tremblant Risk Management conference, the ESSFM Gerzensee 2008 conference, the 2008 Conference on Financial Innovation at Vanderbilt University, the 2008 International Conference on Price, Liquidity, and Credit Risks at Konstanz University; seminars at Mannheim University, Tel Aviv University, University of Cyprus, and Zurich University; and especially Russell Davidson for insightful comments and constructive criticism. We remain responsible for errors and omissions. Constantinides acknowledges financial support from the Center for Research in Security Prices, University of Chicago. Czerwonko and Perrakis acknowledge financial support from the Social Sciences and Humanities Research Council of Canada.

means that the expected utility of a generic risk-averse investor, as defined above, increases when exploiting these violations. This result is independent of any specific preferences, such as mean variance preferences.

Whereas we find a substantial number of violations of the call upper bound, the corresponding put upper bound and call and put lower bounds identify relatively few violations partly because the first two of these bounds are weak, while there do not seem to be many violations of the put lower bound in our data. Since violations of these bounds are too infrequent for statistical inference, we focus on violations of the call upper bound.¹ We note that we can also use our statistical tests as a general method to investigate good buys or good sells and in particular the goodness of alternative methods in finding such opportunities. Finally, we show that neither a simple heuristic for trading options based on observed option price percentiles for buying low and selling high nor an application of the Bernardo and Ledoit (2000) bounds systematically leads to out-of-sample portfolio returns that stochastically dominate portfolio returns that do not trade in options. We conclude that the identification of good buy and good sell options is difficult, yet the Constantinides and Perrakis (2007) bounds achieve this task.

Ample evidence motivates our focus on the class of investors holding the market and the risk-free asset. Surveys report that a large number of U.S. investors follow indexing policies in their investments. Bogle (2005) reports that, in 2004, index funds accounted for about one-third of equity fund cash inflows since 2000 and represented about one-seventh of equity fund assets. The S&P 500 index is not only the most widely quoted market index, but has also been available to investors through exchange traded funds and index futures for several years. Furthermore, the results are robust to the investor's portfolio composition. We show that the writing of good sell options may increase the expected utility of an investor holding a portfolio that includes a wide range of assets in addition to the index and the risk-free asset.

The bounds that identify good sell options are valid for any distribution of the underlying asset, including the empirical ones extracted from past data.² Furthermore, the stochastic dominance statistical tests by Davidson and Duclos (2000, 2006), which we employ to assess the out-of-sample profitability of our trading policy, are valid under minimal technical assumptions about return distributions.

Finally, the results are robust to the estimation of the bounds. Note that we are only using the bound violations as signals for trading. Thus, the assumptions underlying the derivation of the bounds may be violated as long as the signals are sufficiently informative for utility-enhancing trading.

We use the Chicago Mercantile Exchange (CME) database on S&P 500 futures options, from 1983 to 2006, which is clean and spans a long period. Much

¹ In the Internet Appendix, we demonstrate the ability of the lower bounds to identify good buy options by showing that a portfolio with long positions in options bought at artificial prices equal to the lower bound stochastically dominates a portfolio without them.

² In the absence of transaction costs and additional state variables, Oancea and Perrakis (2009) show that the bounds nest the jump diffusion and stochastic volatility option pricing models.

of the earlier empirical work on the mispricing of index options is based on S&P 500 index option data that come from two principal sources: the Berkeley Options Database (from 1986 to 1995), which provides relatively clean transaction prices but misses important events over the past 14 years, such as the 1998 liquidity crisis, the dot-com bubble, and its 2000 burst; and the Option-Metrics database (starting in 1996), which is of uneven quality and contains only end-of-day quotes.

Our tests are nonparametric in the sense that we do not assume any particular distribution for the underlying asset returns. Therefore, our finding of good sell options cannot be attributed to stochastic volatility and jumps in the index price. We use historical data on the underlying S&P 500 index returns to estimate the bounds. We use several empirical estimates of the underlying return distribution, all of them observable at the time the trading policy is implemented. For each of these estimates, we evaluate the corresponding bounds over the period from 1983 to 2006 and then identify the observed S&P 500 futures option prices that violate them. For each violation, we identify the optimal trading policy of a generic investor with and without the option, using the observed path of the underlying asset until option expiration and recognizing realistic trading conditions such as possible early exercise and transaction costs. We identify the profitability of the pair of policies for each observed violation and then conduct several stochastic dominance tests over the entire sample period.

A large body of finance literature addresses the mispricing of options. Rubinstein (1994) and Jackwerth and Rubinstein (1996) observe a steep index smile in the implied volatility of S&P 500 index options that suggests out-of-the-money (OTM) puts are too expensive. Indeed, a common hedge fund policy is to sell OTM puts. Coval and Shumway (2001) find that buying zero beta, at-the-money (ATM) straddles/strangles loses money. Santa-Clara and Saretto (2009) also find that strategies to sell index options are good deals. The results of Aït-Sahalia and Lo (2000), Jackwerth (2000), and Rosenberg and Engle (2002) are suggestive of stochastic dominance, albeit in frictionless markets with a representative agent. The assumption of a representative agent can be justified if the market is complete. Aït-Sahalia and Lo (2000, pp. 25–26) call for extensions that do not rely on complete markets. Our results on stochastic dominance allow for both an incomplete market (thereby removing the requirement that a representative investor exists) and frictions.

Constantinides, Jackwerth, and Perrakis (2009) provide empirical evidence that both European puts and calls on the S&P 500 index violate the corresponding stochastic dominance bounds on European options put forth by Constantinides and Perrakis (2002). Constantinides et al. (2009) estimate the time-series process of the index price, use this process to calculate upper and lower stochastic dominance bounds on option prices, and report the observed violations of the bounds by the option prices. This process is subject to model misspecification and estimation error. Therefore, the bounds are potentially calculated with error. The reported violations do not account for potential error.

We address the above concerns in this paper as follows. The claim of stochastic dominance is not based on the observation of option prices violating the bounds. Observed violation of the bounds only triggers the trading of these options. The claim of stochastic dominance is based on the out-of-sample statistical test of whether the portfolios that incorporate such options stochastically dominate portfolios that do not incorporate them. Even if the assumptions that lead to the theoretical development of the bounds are violated in practice, this does not detract from the finding of stochastic dominance based on the empirical tests, but instead makes the claim of stochastic dominance conservative and implies that the estimated bounds may still be used as identifiers of good sell options.

Unlike the results in Constantinides et al. (2009), our tests of stochastic dominance do allow for error in the realized returns. The statistical tests are based on the 1-month realized distribution of returns of these portfolios. The tests are nonparametric and therefore free from assumptions regarding the return distribution. The tests are not based on the estimated time-series process of the index price and hence are free from estimation error and model misspecification of the time-series process of the index price. The reported findings of stochastic dominance are conservative because potential errors in calculating the bounds result in a trading rule that is less efficient in spotting violations. Finally, we estimate the bounds at time t based only on information available at time t . Therefore, both the calculation of the bounds and the tests of stochastic dominance are truly out of sample.

The paper is organized as follows. In Section I, we present the restrictions on futures option prices imposed by stochastic dominance and discuss the underlying assumptions. In Section II, we describe the data and the empirical design. We present the empirical results in Section III and demonstrate their robustness in Section IV. In Section V, we discuss the implications of our results and conclude. Additional results are reported in the appendices, and in an Internet Appendix available at <http://www.afajof.org/supplements.asp>.

I. Restrictions on Futures Option Prices Imposed by Stochastic Dominance

We summarize the model and assumptions in Constantinides and Perrakis (2007) that lead to the bounds that signify violations of stochastic dominance. We stress that even if these assumptions do not hold in practice, this does not detract from the finding of stochastic dominance based on the empirical tests reported in this paper, but instead makes the claim of stochastic dominance conservative: the tests of stochastic dominance do not depend on the assumptions made in deriving the bounds.

We allow the market to be incomplete and agents to be heterogeneous. We investigate the restrictions on option prices imposed by one particular class of agents that we simply refer to as “traders.” We allow for other agents to participate in the market but this allowance does not invalidate the restrictions on option prices imposed by the traders.

We consider a market with several types of financial assets. First, we assume that traders invest in only two of them, a bond and a stock with the natural interpretation as a market index. Subsequently, we assume that traders can invest in a third asset as well, an American call or put option on the index futures. The bond is risk free and has total return R . The stock has ex dividend stock price S_t at time t and pays cash dividend γS_t , where the dividend yield γ is deterministic. The total return on the stock, $(1 + \gamma)(S_{t+1}/S_t)$, is assumed to be i.i.d. with mean R_S . The call or put option on the index futures has strike K and expiration date T . The underlying futures contract is settled in cash and has maturity T^F , $T^F \geq T$. We assume that the futures price F_t is linked to the stock price by the approximate cost-of-carry relation $F_t = (1 + \gamma)^{-(T^F-t)} R^{T^F-t} S_t + \varepsilon_t$, $t \leq T^F$, $|\varepsilon_t| \leq \bar{\varepsilon}$, where the basis risk ε_t is serially independent and independent of the stock price.

Transfers to and from the cash account (bond trades) do not incur transaction costs. Stock trades decrease the bond account by transaction costs equal to the absolute value of the dollar transaction times the proportional transaction costs rate, k , $0 \leq k < 1$. Transaction costs, exchange fees, and price impact are accounted for in what we refer to as the bid and ask prices of options.

We assume that traders maximize generally heterogeneous, state independent, increasing, and concave utility functions. We further assume that each trader's wealth at the end of each period is weakly monotone increasing in the stock return over the period, as explained in Constantinides and Perrakis (2007).³

We do not make the restrictive assumption that all market agents belong to the class of utility-maximizing traders. Thus, our results are robust and unaffected by the presence in the market of agents with beliefs, endowments, preferences, trading restrictions, and transaction costs schedules that differ from those of the utility-maximizing traders modeled in Constantinides and Perrakis (2007).

A trader enters the market at time zero with x_0 dollars in bonds and y_0 dollars in ex dividend shares of stock. We normalize the stock (or, index price) to y_0 dollars so that the trader holds one share (or, one unit of the index). We consider two scenarios. In the first scenario, the trader may trade the bond and stock but not the options. The trader makes sequential investment decisions at discrete trading dates t ($t = 0, 1, \dots, T'$), where T' , $T' \geq T^F \geq T$, is the finite terminal date. The trader's objective is to maximize expected utility, $E[u_{T'}(W_{T'})]$, where $W_{T'}$ is the trader's net worth at date T' . Utility is assumed to be concave and increasing and defined for both positive and negative terminal worth, but is otherwise left unspecified. We refer to this trader as the *index* (and bond) trader, and denote her maximized expected utility by $V_0^{IT}(x_0, y_0)$.

³ For example, a trader who holds 100 shares of stock and a net short position in 200 call options violates the monotonicity condition, while a trader who holds 200 shares of stock and a net short position in 200 call options satisfies the condition. Essentially, we assume that the traders have a sufficiently large investment in the stock, relative to their net short position in call options (or, net long positions in put options), such that the monotonicity condition is satisfied.

In the second scenario, the trader enters the market at time zero with x_0 dollars in bonds and y_0 dollars in ex dividend shares of stock, but immediately writes one American futures call option with maturity T , $T \leq T^F$, where C are the net cash proceeds from writing the call.⁴ We assume that the trader may not trade the call option thereafter. At each trading date t ($t = 0, 1, \dots, T$), the trader is informed as to whether she has been *assigned* (that is, assigned to act as the counterparty of the holder of a call who exercises the call at that time). If the trader has been assigned, the call position is closed out, the trader pays $F_t - K$ in cash, and the value of the cash account decreases from x_t to $x_t - (F_t - K)$. The trader makes sequential investment decisions with the objective of maximizing her expected utility, $E[u_{T'}(W_{T'})]$. We refer to this trader as the *option* (plus index and bond) trader, and denote her maximized expected utility by $V_0^{OT}(x_0 + C, y_0)$.

For a given pair (x_0, y_0) , we define the reservation write price of a call as the value of C such that $V_0^{OT}(x_0 + C, y_0) = V_0^{IT}(x_0, y_0)$. The interpretation of \bar{C} is the write price of the call at which the trader with initial endowment (x_0, y_0) is indifferent between writing the call or not. Constantinides and Perrakis (2007) state a tight upper bound on the reservation write price of an American futures call option that is independent of the trader's utility function and initial endowment and independent of the early exercise policy on the calls

$$\bar{C}(F_t, S_t, t) = \frac{1+k}{1-k} \max[N(S_t, t), F_t - K], \quad t \leq T. \quad (1)$$

The function $N(S, t)$ is defined as follows:

$$\begin{aligned} N(S, t) &= (R_S)^{-1} E[\max\{(1+\gamma)^{-(T^F-t-1)} R^{T^F-t-1} S_{t+1} \\ &\quad + \bar{\varepsilon} - K, N(S_{t+1}, t+1)\} | S_t = S], \quad t \leq T-1 \\ &= 0, \quad t = T. \end{aligned} \quad (2)$$

The economic interpretation of the call upper bound is as follows. If we observe a call bid price above the reservation write price, C , then any trader (as defined in this paper) can increase her expected utility by writing the call.

Transaction costs on the index have only a small effect on the upper bound. Without transaction costs on the index, the upper bound is $\max[N(S_t, t), F_t - K]$; with transaction costs on the index, the upper bound increases merely by the multiplicative factor $(1+k)/(1-k)$. The reason is that this particular bound is based on a comparison of the utility of an index trader to the utility of an option trader. Both traders follow the trading policy that is optimal for the index trader but is generally suboptimal for the option trader. This policy incurs very low transaction costs because trades are infrequent, as shown in Constantinides (1986).

If we further assume that the trader can buy a call at price $\bar{C}(F_t, S_t, t)$ or less and trade the futures and do so costlessly, we obtain the following put upper

⁴ The reservation write price of a call is derived from the perspective of a trader who is marginal in the index, the bond, and only one type of call or put option at a time. Therefore, these bounds allow for the possibility that the options market is segmented.

bound.⁵

$$P(F_t, S_t, t) = C(F_t, S_t, t) - R^{-(T-t)}F_t + K, \quad t \leq T, \quad (3)$$

which has a similar interpretation. Constantinides and Perrakis (2007) derive lower bounds on the reservation purchase price of American call and put options on futures. We do not state these lower bounds here because the observed frequency of their violation is too low for statistical inference.

II. Data Description and Methodology

In this section, we describe our data on index, futures, and option prices. We explain how we calibrate a tree of the daily index return and use it to calculate the option bounds. We describe the construction of the portfolio of the index trader (IT) and of the option trader (OT). Finally, we explain our empirical methodology of comparing the performance of the IT and OT portfolios in terms of both their means and the criterion of second-order stochastic dominance.

A. Data Description and Estimation

We obtain the time-stamped quotes of the 30-calendar-day S&P 500 futures options and the underlying nearest-to-maturity futures for the period from February 1983 to July 2006 from the CME tapes. This results in 247 sampling dates. We obtain the interest rate as the 3-month T-bill rate from the Federal Reserve Statistical Release H.15. The data sources are described in further detail in Appendix A.

For the daily index return distribution, we use the historical sample of logarithmic returns from January 1928 to January 1983. However, when looking forward for each of our 247 option sampling dates, we adjust the first four moments of the index return distribution in various ways, which we now describe in detail. We set the mean logarithmic index return at 4% plus the observed 3-month T-bill rate instead of estimating the mean index return from the data in order to mitigate statistical problems in estimating the mean. We implement this by adding a constant to the observed logarithmic index returns so that their sample mean equals the above target.

We estimate both the unconditional and conditional volatility of the index returns. We estimate the unconditional volatility as the sample standard deviation over the period January 1928 to January 1983.

We estimate the conditional volatility in three different ways. First, we estimate the conditional volatility as the sample standard deviation over the preceding 90 trading days. We also estimate the conditional volatility over

⁵ We prove equation (3) by noting that an investor achieves an arbitrage profit by buying a call at $\bar{C}(F_t, S_t, t)$; writing a put at P , $P > \bar{P}(F_t, S_t, t)$; selling one future; and lending $K - R^{-(T-t)}F_t$. In the proof, we ignore the daily marking to market on the futures until the exercise of the put or the options' maturity, whichever comes first. This matters little because the investor has a large investment in the bond, which suffices to cover margin calls.

Table I
Prediction Error of Monthly Volatility, 1983–2006

The errors are defined as the difference between the monthly volatility and the volatility predicted by a given mode. The unconditional volatility is the sample standard deviation over the period January 1928 to January 1983. The 90-day volatility is the sample standard deviation over the preceding 90 trading days. The adjusted IV is the ATM IV on the preceding day, adjusted by the mean prediction error for all dates preceding the given date, where we drop from the preceding days all 21 precrash observations. The EGARCH volatility is the volatility using EGARCH coefficients estimated for S&P 500 daily returns over January 1928 to January 1983 and applied to residuals observed over the 90 days preceding each sample date to form projections of the volatility realized up to the option expiration date.

Prediction Mode	Mean	Median	St. Dev.	Skew.	Ex. Kurt.
Unconditional	0.0429	0.0649	0.0680	-1.7300	3.8296
90 day	0.0095	0.0076	0.0595	0.2687	5.2490
Adjusted IV	-0.0005	0.0002	0.0496	-0.2625	3.4680
EGARCH	0.0177	0.0185	0.0531	0.0936	7.8302

the preceding 360 days. The results remain essentially unchanged. Second, we estimate the conditional volatility as the ATM implied volatility (IV) on the preceding day, adjusted by the mean prediction error for all dates preceding the given date (typically some 3%).⁶ Finally, we estimate the conditional volatility as the Nelson (1991) EGARCH (1, 1) model volatility using EGARCH coefficients estimated for S&P 500 daily returns over January 1928 to January 1983 applied to residuals observed over the 90 days preceding each sample date to form projections of the volatility realized until the option expiry date.⁷ We estimate the third and fourth moments of the index return as their sample counterparts over the preceding 90 days.

In Table I, we report statistics of the prediction error of the above volatility estimates. The best overall predictor is the adjusted ATM IV and the second-best predictor is the 90-day historical volatility.

B. Calibration of the Index Return Tree and Calculation of the Option Bounds

We model the path of the daily index return up to the option expiration on a T -step tree, where T is the number of trading days in the particular month. The tree is recombining with m branches emanating from each node. Each month we calibrate the tree by choosing the number of branches, spacing,

⁶ We start with the 22nd month. We use the holdout sample of the first 21 months to estimate the mean adjustment error and adjust the IV of the 22nd month. We use the holdout sample of the first 22 months to estimate the mean adjustment error and adjust the IV of the 23rd month; and so on.

⁷ We form the volatility projections by iterating from day $t + 1$ till the option maturity T , as explained in Baillie and Bollerslev (1992). We use as inputs the past 90 day residuals and the model coefficients estimated in the presample period. In the final step, we sum up the forecasted squared residuals to derive the variance forecast for a given period. The estimated model coefficients were as follows: κ -0.10451, ARCH (1) 0.16620, GARCH(1) 0.98799, leverage -0.05969.

and transition probabilities at each node to match the first four moments of the daily index return distribution, as described in the Internet Appendix. We numerically calculate the bounds in equations (1) to (3) by iterating backward on the calibrated tree.

C. Portfolio Construction and Rebalancing

For each monthly stock return path, we employ the following rebalancing policies. For the index trader (who manages a portfolio of the index and the risk-free asset in the presence of transaction costs), we employ the optimal trading policy derived in Constantinides (1986) and extended in Perrakis and Czerwonko (2006) to allow for a dividend yield on the stock. Essentially, this policy consists of trading only to confine the ratio of the index value to the bond value, y_t/x_t , within a no-transactions region, defined by lower and upper boundaries.

We derive these lower and upper boundaries for the following parameter values: one-way transaction cost rate on the index, 0.5%; annual return volatility of the index equal to the sample volatility over 1928 to 1983, 0.1856; interest rate equal to the observed 3-month T-bill rate; risk premium, 4%; and constant relative risk aversion coefficient, 2. For this set of parameters, the lower and upper boundaries are $y_0/x_0 = 1.2026$ and 1.5259 , respectively. At the beginning of each month and before the trader trades in options, we set $x_0 = 73,300$ and $y_0 = 100,000$, which corresponds to the midpoint of the no-transactions region, $y_0/x_0 = 1.3642$. We normalize the index price to y_0 dollars so that the trader holds one unit of the index. In our empirical work, we verify the robustness of our results to these parameter values and the initial portfolio composition.

For the option trader (who manages a portfolio of the option, the index, and the risk-free asset in the presence of transaction costs), we set x_0 and y_0 to the same values as for the index trader. The option trader writes or buys one call or one put on the index futures. We normalize the size of a futures contract to be on one unit of the index, and we normalize the size of the futures option to be on one futures contract.

The portfolio composition changes depending on the assumed position in futures options, as explained in Appendix B. We employ the trading policy that is optimal for the index trader but is generally suboptimal for the option trader. Recall that the goal is to demonstrate that there exist profitable investment opportunities for the option trader. Given this goal, it suffices to show that there exist profitable investment opportunities for the option trader even though the option trader follows a generally suboptimal policy in trading the index.

We focus on the case in which the basis risk bound, $\bar{\varepsilon}$, is 0.5% of the index price. Over the years from 1990 to 2002, 95% of all observations have basis risk less than 0.5% of the index price. In Section IV, we argue that our empirical results are robust to the basis risk bound.

D. Empirical Methodology

For each of our methods of estimating the bounds, we obtain 247 monthly portfolio returns for the index trader and the option trader. Our goal is to test

whether the portfolio profitabilities of the index and option traders are statistically different in the months in which we observe violations of the bounds.

We apply the criterion of second-order stochastic dominance (SSD), which states that the dominating portfolio is preferred by any risk-averse trader, independent of distributional assumptions, such as normality, and preference assumptions, such as quadratic utility. Formally, the OT portfolio stochastically dominates the IT portfolio if, for every z in the joint support of their respective distributions, the following holds:

$$D_{IT}^2(z) - D_{OT}^2(z) \geq 0, \quad (4)$$

with strict inequality for at least one value of z , where

$$D_J^2(z) = \int_{\underline{z}}^z (z - x) dF_J(x), \quad (5)$$

$J = OT, IT$, $F_J(x)$, is the cumulative distribution function of the portfolio return, and \underline{z} is the lower bound of the common support.

First, we test the null hypothesis $H_0 : IT \succ_2 OT$ against the alternative that either $OT \succ_2 IT$ or neither one of the two distributions dominates the other. Hence, rejection of the null hypothesis fails to rank the two distributions. We also test the converse null hypothesis $H_0 : OT \succ_2 IT$ against the alternative that either $IT \succ_2 OT$ or neither one of the two distributions dominates the other. For these hypotheses, we report the results of the test proposed by Davidson and Duclos (2000) (DD (2000)), described in Appendix C. The test requires that returns be serially uncorrelated, an assumption that holds well in all our return series: the first-order serial correlation ranges from -0.0267 to 0.0964 and is statistically insignificant.

Second, we test the null hypothesis $H_0 : OT \not\succeq_2 IT$, which states that the option trader's portfolio return does not stochastically dominate the index trader's portfolio return, against the alternative hypothesis $H_A : OT \succ_2 IT$, which states that the option trader's portfolio return stochastically dominates the index trader's portfolio return. Rejection of this hypothesis means that the option trader's portfolio return stochastically dominates the index trader's portfolio return. Similarly, we test the converse null hypothesis $H_0 : IT \not\succeq_2 OT$ against the alternative hypothesis $H_A : IT \succ_2 OT$. For these hypotheses, we report the results of the test proposed by Davidson and Duclos (2006) (DD (2006)), described in Appendix C. We use the algorithm developed by Davidson (2007). Again, the test requires that returns be serially uncorrelated, an assumption that holds well in all our return series. Nolte (2008) investigates the power of the DD (2006) test when there are GARCH effects and finds that the test performs well.

The power of the DD (2006) test is low, unless one trims the tails of the paired outcomes. We therefore trim 10% of the paired outcomes in the left tail of our sample distributions, which affects both the IT and the OT returns similarly and is therefore innocuous. The trimming of the right tail of the distribution presents a problem. Without any trimming on the right tail, the test

has low power. Trimming on the right, on the other hand, may bias our test toward rejection of the null, because IT tends to produce superior results to OT when the return of the underlying asset is high. For this reason, in our tables we present results with 0%, 5%, and 10% trimming on the right. In the Internet Appendix, we test the effects of such trimming on simulated data that mirror our sample. Our simulations show that the test is, if anything, conservative in rejecting the false null. To facilitate interpretation, we perform all our statistical tests on annualized arithmetic returns on the wealth of OT and IT investors.⁸

We choose the tests DD (2000) and DD (2006) because, unlike several alternatives, they apply to correlated samples and are more powerful than other well-known tests. For further discussion, see Tse and Zhang (2003).

III. Empirical Results

In Section III.A, we describe the pattern of observed violations for the bounds with respect to the degree of moneyness. In Section III.B, we present the main empirical results. We compare the portfolio return of an option trader who writes good sell calls or puts at their bid price with the portfolio return of an index trader who does not trade in the options over the period from 1983 to 2006. We find that the return of an option writer stochastically dominates the index trader's return, net of transaction costs and the bid-ask spread. Whereas we find a substantial number of violations of the upper bounds, we find relatively few violations of the lower bounds.

A. The Pattern of Violations

In Figure 1, we plot the four bounds for 1-month options for May 22, 1996, expressed in terms of the implied volatility, as a function of the moneyness, K/F . We set $\sigma = 20\%$ and $\bar{\varepsilon} = 0$. The figure also displays the 95% confidence interval, derived by bootstrapping the 90-day distribution. The call upper bound is tighter than the put upper bound and both bounds are downward sloping. The put lower bound is tighter than the call lower bound. The put lower bound is downward sloping but the call lower bound is not.

In Figure 2, we display the time pattern of actual violations of the call upper bound. The crosses display the violations of the call upper bound for the period February 1983 to July 2006. For the adjusted IV distributions, the first 21 dates are not in the sample because they are needed to obtain the adjustment. The solid lines are the natural logarithm of the S&P 500 index, the VIX index, and the T-bill rate. For all different ways of estimating volatility, we observe violations after significant down moves in the index, when we expect the implied volatility to be high.

⁸ We annualize returns since times to maturity vary from 28 to 31 days in our sample. Since transaction costs are present in our economy, we derive returns for the liquidation of the risky asset under the assumed one-way transaction costs rate of 0.5%.

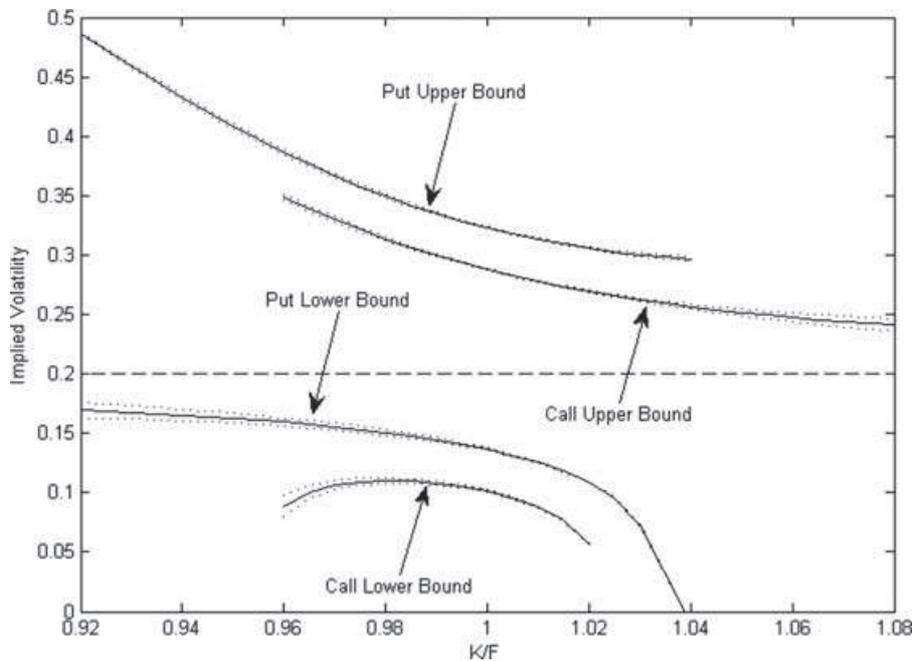


Figure 1. Illustration of upper and lower bounds on call and put options. Bounds are derived for $\sigma = 0.20$ imposed on a 90-day past distribution of daily S&P 500 returns for May 22, 1996. 95% upper and lower confidence intervals represented by dotted lines are derived by bootstrapping the 90-day distribution. The results exemplify the dependence of the bounds on the third and fourth moments of the distribution because the width in the confidence intervals is determined solely by varying the skewness and kurtosis, that is, the bootstrap changes only these quantities.

In Table II, the violations are shown as a proportion of the quotes in each moneyness range. We note that, for all methods of estimating the bounds, the proportion of violations is large. For the moneyness range 1.03 to 1.08, a large proportion of the available quotes violates the corresponding bound for all estimation methods.

Table III shows the violations in each moneyness range as a proportion of the total number of quotes across the whole range of moneyness. The largest number of violations, as a proportion of the total number of quotes, is found in the 1.01 to 1.03 moneyness range and not in the 1.03 to 1.08 range because there are relatively few quotes in the latter range. For all estimation methods, a majority of the identified violations are in the liquid range, 0.99 to 1.03. The exchange regulations specify that the minimum number of available contracts must be at least 20 for each quote. Our data also show that the average size of the violation is between 5% and 56% of the upper bound for most methods of estimating the bounds as we move from the 0.96 to 0.99 moneyness range to the 1.03 to 1.08 moneyness range. In the stochastic dominance tests, the power of the tests depends, by construction, on the proportion of months with observed violations.

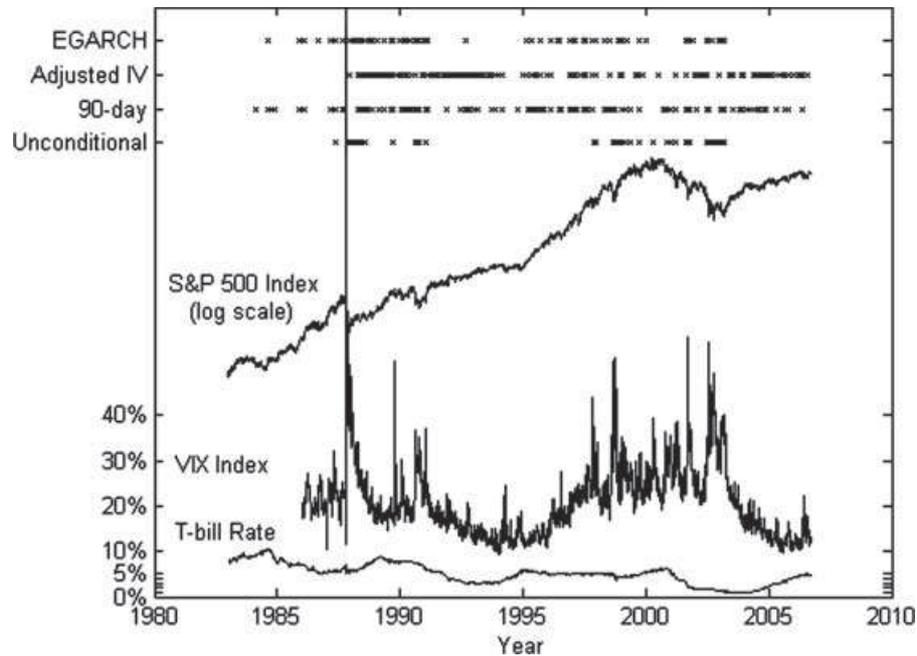


Figure 2. Time distribution of observed violations. The crosses display the violations of the call upper bound for the period February 1983 to July 2006. For the adjusted IV distribution, the first 21 dates are not in the sample. To facilitate presentation, the S&P Index was transformed to a logarithmic scale. The inception date of the VIX index was on February 4, 1986. The value for VIX just prior to the October 1987 crash was 170% and is trimmed to facilitate presentation.

Table II
Percentage of Call Quotes with Violations of the Upper Bound

The table displays the percentages of call bids violating the call upper bound out of all bid quotes observed in each respective moneyness bracket.

Volatility Prediction Mode	Moneyness (K/F) Range			
	0.96–0.99	0.99–1.01	1.01–1.03	1.03–1.08
Unconditional	3.3	5.8	8.6	20.6
90 day	4.8	16.1	32.2	45.4
IV adjusted	4.1	15.2	30.5	30.3
EGARCH	2.3	6.5	14.1	24.3

We further investigate how the incidence of calls violating the upper bound relates to characteristics of the options (moneyness, ATM implied volatility, volume of trade, and put-call ratio), the index (return, momentum, presence of jumps between trading dates, dividend yield, skew, and trading volume), the term structure, and the default spread. For each of the four volatility prediction methods, we sort the sample of calls into terciles and report the average

Table III
Percentage of Call Quotes with Violations of the Upper Bound Out of the Total Number of Quotes

The table displays the percentages of call bids violating the call upper bound in each respective moneyness bracket out of the total number of observed bid quotes.

Volatility Prediction Mode	#Months with Viol. (#Months)	Moneyness (K/F) Range				
		0.96–0.99	0.99–1.01	1.01–1.03	1.03–1.08	All
Unconditional	43 (247)	0.4	1.7	3.5	3.8	9.4
90 day	100 (247)	0.5	4.8	13.2	8.4	26.9
IV adjusted	120 (226)	0.3	4.6	12.9	5.9	23.7
EGARCH	65 (247)	0.3	1.9	5.8	4.5	12.5

fraction of violations in the top and bottom terciles. The results are reported in Table IV. The incidence of violations is higher when the Moody Baa-Aaa default spread is high, when the futures open interest is high, and also when momentum is high—situations that suggest investor uncertainty is rather high. This nervousness might well transmit to the option market and increase the likelihood of observing violating options. For all other characteristics, we do not find a consistent pattern of violations because the incidence of violations depends on the prediction method of the volatility as an input to the derivation of the bounds.

B. Empirical Evidence on Stochastic Dominance

We apply our statistical tests to all months in the sample, even though there are months in which the OT trader does not trade in options and the returns in these months are identical for the OT and IT portfolios, thereby making it harder for us to establish profitable (utility-improving) trading opportunities. In Table V, Panels A and B, we present the cases of call and put bid prices violating their upper bound, when we set the basis risk bound at 0.5% of the index price. We find a higher frequency of violations of the call upper bound than of the upper put bound because the call upper bound is tighter than the put upper bound, as we observed in Figure 1.⁹

In our first test of stochastic dominance, we consider the hypothesis $H_0 : OT \succ_2 IT$, which states that the option trader's return dominates the index trader's return. We apply the DD (2000) test and obtain p -values that exceed 10% for both the call upper bound and the put upper bound.

In our second test, we consider the hypothesis $H_0 : IT \succ_2 OT$, which states that the index trader's return dominates the option trader's return. Again, we apply the DD (2000) test. In Table V, Panel A, the p -values are lower than 1%

⁹ The theory in Constantinides and Perrakis (2007) has not produced a tight upper bound for put futures options. The bound in (3) is a weak filter extracted from the combination of the call upper bound and put-call parity.

Table IV
Pattern of Bounds Violations

The classified variable is the ratio of the quotes in violation of the call upper bound to the overall number of call bid quotes observed in a given cross-section. The number of cross-sections is 247 for all volatility prediction modes except for Adjusted IV, where it is 226. The overall means (std) of the classified variable for each volatility prediction mode are: unconditional 0.13 (0.32), 90 day 0.22 (0.35), Adjusted IV 0.24 (0.32), and EGARCH 0.13 (0.29). Option characteristics: mean K/F is the average moneyness of call bid quotes, ATM IV is the average implied volatility of several bid and ask quotes for calls and puts closely bracketing a K/F ratio of 1, call and put volume (open interest) is the natural logarithm of the respective volume (open interest) recorded on the previous day net of the natural logarithm of the average of the respective volume (open interest) over the past 90 calendar days, put-call ratio is the put volume divided by the call volume, and put-call ratio open interest is the put open interest divided by the call open interest with both values recorded on the previous day. Index characteristics: return is the S&P 500 index excess return over the previous month, momentum is the ratio of the previous-day S&P 500 index level to its average level over the past year, negative (positive) jumps represent the number of S&P 500 daily returns lower than -4% (greater than 4%) recorded over the past 30 calendar days (15 (16) occurrences for negative (positive) jumps in total), yield is the ratio of dollar dividends over the past year to the previous-day S&P 500 index level, left (right) skew is the implied volatility for the K/F ratio of 0.96 net of ATM IV (ATM IV net of the implied volatility for the K/F ratio of 1.04), and futures volume (open interest) is the natural logarithm of the futures volume (open interest) recorded on the previous day net of the natural logarithm of the average of the futures volume (open interest) over the past 90 calendar days. Interest characteristics: interest rate is the 3-month T-bill rate, yield slope is the difference between the yield of 10- and 1-year to maturity CRSP indices for government bonds, and Moody's Baa – Aaa is the difference between the baskets of Baa- and Aaa-ranked corporate bonds.

	Volatility Prediction Mode							
	Unconditional		90 day		Adjusted IV		EGARCH	
	Bottom Tercile	Top Tercile	Bottom Tercile	Top Tercile	Bottom Tercile	Top Tercile	Bottom Tercile	Top Tercile
Option Characteristics								
Mean K/F	0.042	0.278	0.087	0.331	0.248	0.185	0.074	0.194
ATM IV	0.000	0.385	0.100	0.373	0.243	0.182	0.028	0.275
Call volume	0.109	0.178	0.117	0.298	0.303	0.205	0.082	0.186
Put volume	0.110	0.137	0.161	0.263	0.288	0.227	0.127	0.097
OTM put volume	0.103	0.152	0.188	0.241	0.309	0.230	0.105	0.078
Call open interest	0.130	0.186	0.172	0.271	0.255	0.250	0.109	0.172
Put open interest	0.139	0.152	0.180	0.274	0.267	0.246	0.090	0.162
OTM put open interest	0.138	0.167	0.221	0.185	0.240	0.249	0.102	0.198
Put call ratio	0.150	0.124	0.167	0.225	0.234	0.234	0.126	0.066
Put call ratio, open interest	0.228	0.079	0.226	0.197	0.225	0.275	0.211	0.064
Index Characteristics								
Return	0.244	0.093	0.357	0.156	0.226	0.261	0.146	0.192
Momentum	0.045	0.312	0.242	0.286	0.184	0.229	0.099	0.223
Negative jumps ^a	0.095	0.784	0.211	0.302	0.248	0.115	0.119	0.343
Positive jumps ^a	0.109	0.639	0.224	0.010	0.251	0.006	0.135	0.000
Yield	0.172	0.116	0.219	0.208	0.186	0.381	0.126	0.145
Left skew	0.121	0.189	0.166	0.255	0.341	0.117	0.103	0.159
Right skew	0.042	0.276	0.160	0.261	0.357	0.150	0.118	0.189
Futures volume	0.090	0.184	0.155	0.301	0.267	0.252	0.118	0.151
Futures open interest	0.066	0.212	0.159	0.290	0.218	0.284	0.098	0.193

(continued)

Table IV—Continued

	Volatility Prediction Mode							
	Unconditional		90 day		Adjusted IV		EGARCH	
	Bottom Tercile	Top Tercile	Bottom Tercile	Top Tercile	Bottom Tercile	Top Tercile	Bottom Tercile	Top Tercile
Interest Characteristics								
Interest rate	0.150	0.128	0.205	0.231	0.292	0.281	0.124	0.149
Yield slope	0.107	0.134	0.191	0.178	0.239	0.292	0.098	0.101
Moody's Baa – Aaa	0.033	0.274	0.205	0.258	0.134	0.317	0.104	0.208

^aInstead of terciles, the classified variable was split for dates without and with jumps in the preceding 30 calendar days.

Table V
Returns of Call Trader and Index Trader

The equally weighted average of all violating options equivalent to one option per share traded at each date. The symbols * and ** denote a difference in sample means of the *OT* and *IT* traders significant at the 5% and 1% levels in a one-sided bootstrap test with 9,999 trials. Maximal *t*-statistics for the Davidson and Duclos (2000) test are compared to critical values of the Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for nominal levels of 1%, 5%, and 10% with $k = 20$ and $\nu = \infty$. The *p*-values for $H_0 : OT \succ_2 IT$ are greater than 10%, the highest nominal level available in the Stoline and Ury (1979) tables. The *p*-values for the Davidson and Duclos (2006) test are based on 999 bootstrap trials. The *p*-values for $H_0 : IT \not\prec_2 OT$ are equal to 1.

Volatility Prediction Mode	# Months with Viol. (# Months)	$\hat{\mu}_{OT} - \hat{\mu}_{IT}$ (Annualized)	DD (2000) p -Value $H_0 : IT \succ_2 OT$	DD (2006) p -Value $H_0 : OT \not\prec_2 IT$ 10% Trimming in Left Tail, Trimming in Right Tail as Below:		
				No Trimming	5% Trimming	10% Trimming
Panel A: Call Upper Bound						
Unconditional	43 (247)	0.0031	<0.01	0.244	0.024	0.000
90 day	100 (247)	0.0043	<0.01	0.166	0.007	0.002
Adjusted IV	120 (226)	0.0066*	<0.01	0.119	0.029	0.000
EGARCH	65 (247)	0.0062**	<0.01	0.079	0.000	0.000
Panel B: Put Upper Bound						
Unconditional	23 (247)	0.0009	>0.1	0.399	0.203	0.154
90 day	16 (247)	-0.0008	>0.1	1	1	1
Adjusted IV	4 (226)	n/a	n/a	n/a	n/a	n/a
EGARCH	9 (247)	n/a	n/a	n/a	n/a	n/a

and the hypothesis is rejected when the option trader writes good sell calls. In the first two rows of Panel B, the *p*-values exceed 10% but that is largely because there are very few months in which we observe violations of the put upper bound.

In our third test, we consider the hypothesis $H_0 : IT \not\prec_2 OT$, which states that either the option trader's return dominates the index trader's return or neither return dominates the other. We apply the DD (2006) test and obtain p -values of one for both the call upper bound and the put upper bound.

Finally, we consider the hypothesis $H_0 : OT \not\prec_2 IT$, which states that either the index trader's return dominates the option trader's return or neither return dominates the other. As we explained earlier, the power of the DD (2006) test is low, unless we trim the tails of the paired outcomes. Therefore, we trim 10% of the paired outcomes in the left tail of our sample distributions. Without trimming on the right tail, the test has low power and the p -values are high in the two panels of Table V. Without trimming, we reject at the 10% level for one case in Panel A. In Panel A, with either 5% or 10% trimming on the right tail, the null is rejected. In Panel B, the results are inconclusive largely because there are few months in which we observe violations of the put upper bound.

In Table V, the reported annualized return difference between the portfolios of the OT and IT traders appears low because the written call is a small fraction of the dollar investment in the portfolio: for each unit of the index held in the portfolio, the OT trader writes only one call. In fact, the identified good sell call options provide average *monthly* returns of 12.3% to 43.2% to an investor who writes them at their bid price. The range 12.3% to 43.2% corresponds to the four different ways of estimating the volatility; in three out of these four cases, the average returns are statistically significant according to the bootstrap test with 9,999 trials.

We repeat our run for the subset of OTM good sell calls with little change in the results. To put these results in perspective, the average monthly returns of all call options written at their bid price is 7.7% and this return is not statistically significantly different from zero.

A second reason for the rather low portfolio returns is that we report the average return difference of all months, including the months in which there are no violations of the bounds and no option writing. For example, suppose that there are violations in n months and no violations in m months; further suppose that the option writing over the n months increases the annualized return in these months by $x\%$ on average. In the table, we report the annualized return difference as $xn/(n+m)\%$ and not as $x\%$. As shown in the table, even the largest number of dates with violations is still less than half the total number of dates in our sample. Therefore, the excess returns attributable to the options strategies should be multiplied by a factor greater than two. The economic significance can be further raised by employing portfolios with more than one call option per unit of the underlying future. We defer the presentation of these results until Section IV.B and Table VII, which detail the robustness of our approach.

Overall, the results in Table V show that the relatively large number of violations of the call upper bound by call bid prices leads to a trading policy where the option trader's return stochastically dominates the index trader's

Table VI
Robustness Tests of the Initial Portfolio Composition

The equally weighted average of all violating options equivalent to one option per share was traded at each date. The symbols * and ** denote a difference in sample means of the *OT* and *IT* traders significant at the 5% and 1% levels in a one-sided bootstrap test with 9,999 trials. Maximal *t*-statistics for the Davidson and Duclos (2000) test are compared to critical values of the Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for nominal levels of 1%, 5%, and 10% with $k = 20$ and $v = \infty$. The *p*-values for $H_0 : OT >_2 IT$ are greater than 10%, the highest nominal level available in the Stoline and Ury (1979) tables. The *p*-values for the Davidson and Duclos (2006) test are based on 999 bootstrap trials. The *p*-values for $H_0 : IT \not>_2 OT$ are equal to 1. MSCI denotes Morgan Stanley Capital Index (in Panel E, data are absent for the first four dates in our sample). REIT denotes real estate investment fund. HML and SMB denote Fama and French (1993) factors, that is HML denotes a portfolio long in two value portfolios and short in two growth portfolios and SMB denotes a portfolio long in three small cap portfolios and short in three large cap portfolios. See Fama and French (1993) for details.

Volatility Prediction Mode	# Months with Viol. (# Months)	$\hat{\mu}_{OT} - \hat{\mu}_{IT}$ (Annualized)	DD (2000) <i>p</i> -Value $H_0 : IT >_2 OT$	DD (2006) <i>p</i> -Value $H_0 : OT \not>_2 IT$ 10% Trimming in Left Tail, Trimming in Right Tail as Below:		
				No Trimming	5% Trimming	10% Trimming
Panel A: 100% CRSP Value-Weighted Index in Stock Account						
Unconditional	43 (247)	0.0031	<0.01	0.244	0.001	0.001
90 day	100 (247)	0.0043	<0.01	0.174	0.007	0.000
Adjusted IV	120 (226)	0.0066*	<0.01	0.124	0.014	0.000
EGARCH	65 (247)	0.0062**	<0.01	0.084	0.000	0.000
Panel B: 100% MSCI in Stock Account						
Unconditional	43 (247)	0.0031	<0.05	0.244	0.031	0.000
90 day	100 (247)	0.0043	<0.01	0.174	0.015	0.003
Adjusted IV	120 (226)	0.0066*	<0.01	0.124	0.019	0.010
EGARCH	65 (247)	0.0062**	<0.01	0.084	0.003	0.000
Panel C: 60% of S&P 500 and 40% of CRSP Zimen REIT Value-Weighted Index in Stock Account						
Unconditional	43 (247)	0.0031	<0.05	0.244	0.038	0.002
90 day	100 (247)	0.0043	<0.01	0.175	0.018	0.000
Adjusted IV	120 (226)	0.0066*	<0.01	0.124	0.050	0.006
EGARCH	65 (247)	0.0062**	<0.01	0.083	0.000	0.000
Panel D: 60% of S&P 500, 20% of HML, and 20% of SMB in Stock Account						
Unconditional	43 (247)	0.0031	<0.05	0.244	0.126	0.038
90 day	100 (247)	0.0043	<0.01	0.149	0.056	0.012
Adjusted IV	120 (226)	0.0066*	<0.01	0.133	0.052	0.000
EGARCH	65 (247)	0.0062**	<0.01	0.072	0.006	0.000
Panel E: 60% of S&P 500 and 40% of Barclays Aggregate U.S. Bond Index in Stock Account						
Unconditional	33 (212)	0.0026	<0.05	0.288	0.100	0.031
90 day	84 (212)	0.0056*	<0.01	0.150	0.018	0.000
Adjusted IV	94 (212)	0.0060	<0.01	0.162	0.094	0.000
EGARCH	48 (212)	0.0065*	<0.01	0.087	0.003	0.000
Panel F: 60% of S&P 500 and 40% of Moody's Commodity Index in Stock Account						
Unconditional	43 (247)	0.0031	<0.05	0.244	0.106	0.042
90 day	100 (247)	0.0043	<0.01	0.175	0.034	0.001
Adjusted IV	120 (226)	0.0066*	<0.01	0.124	0.061	0.021
EGARCH	65 (247)	0.0062**	<0.01	0.085	0.000	0.000

return. Furthermore, the observed violations of the call upper bound are economically significant: we find that stochastic dominance evades rejection only after we artificially decrease the price of the written calls by 10% to 15%.

The statistical significance of violations of the put upper bound by put bid prices is weak because there are few months in which we observe violations of the put upper bound. Since several earlier studies such as Bondarenko (2003) and Driessen and Maenhout (2007) identify short puts as highly attractive investments, we further investigate the issue of the existence of good sell put options by testing the policy of shorting straddles and strangles triggered by observing call options that violate their upper bounds at the same or similar strike price. The results, reported in the Internet Appendix, are consistent with put mispricing documented in these earlier studies.

IV. Robustness Tests

We address the robustness of the results to the composition of the trader's initial portfolio (Section IV.A), the number of written options (Section IV.B), the rebalancing of the portfolio (Section IV.C), and the determination of the bounds (Section IV.D). We find that the results presented in Table V are robust. In Section IV.E, we demonstrate that the option upper bounds produce trading results superior to both a naïve trading policy and a policy based on the Bernardo and Ledoit (2000) bounds.

A. Robustness to the Initial Portfolio Composition

We test the sensitivity of the results to the assumption that the index trader does not hold any assets other than the risk-free asset and the index. We vary the composition of the IT portfolio by adding a broad range of risky assets and investigate whether the trader improves her utility by writing the earlier-identified good sell calls. The bounds that we rely upon are no longer strictly valid because they are derived under the assumption that the index trader does not hold any additional assets beyond the index and the risk-free asset. Nevertheless, we rely on these bounds to find good sell calls, thereby making it more challenging to find a profitable OT strategy.

In Table VI, we consider several cases where the IT portfolio differs from a portfolio invested in the S&P 500 index and the risk-free asset either by replacing the investment in the S&P 500 index with an investment in different indices (Panels A to C) or by varying the percentage investment in the index and broadening the portfolio to include the "high minus low" (HML) and "small minus big" (SMB) excess returns of Fama and French (1993) and a real estate fund (Panels D to F); in the Internet Appendix, we also consider cases of open option positions. In all cases, we find that the portfolio return of the option trader who writes good sell calls stochastically dominates the return of the modified index trader.

Table VII
Robustness Tests of the Number of Written Calls

The equally weighted average of all violating options was traded at each date. The symbols * and ** denote a difference in sample means of the *OT* and *IT* traders significant at the 5% and 1% levels in a one-sided bootstrap test with 9,999 trials. Maximal *t*-statistics for the Davidson and Duclos (2000) test are compared to critical values of the Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for nominal levels of 1%, 5%, and 10% with $k = 20$ and $\nu = \infty$. The *p*-values for $H_0 : OT >_2 IT$ are greater than 10%, the highest nominal level available in the Stoline and Ury (1979) tables. The *p*-values for the Davidson and Duclos (2006) test are based on 999 bootstrap trials. The *p*-values for $H_0 : IT \not>_2 OT$ are equal to 1.

Volatility Prediction Mode	# Months with Viol. (# Months)	$\hat{\mu}_{OT} - \hat{\mu}_{IT}$ (Annualized)	DD (2000) p -Value $H_0 : IT >_2 OT$	DD (2006) p -Value $H_0 : OT \not>_2 IT$ 10% Trimming in Left Tail, Trimming in Right Tail as Below:		
				No Trimming	5% Trimming	10% Trimming
Panel A: One and Half Calls per Unit of Index						
Unconditional	43 (247)	0.0046	<0.01	0.244	0.089	0.000
90 day	100 (247)	0.0064	<0.01	0.166	0.055	0.006
Adjusted IV	120 (226)	0.0099*	<0.01	0.122	0.053	0.000
EGARCH	65 (247)	0.0092**	<0.01	0.097	0.018	0.000
Panel B: Two Calls per Unit of Index						
Unconditional	43 (247)	0.0062	<0.01	0.242	0.125	0.026
90 day	100 (247)	0.0086	<0.01	0.166	0.084	0.025
Adjusted IV	120 (226)	0.0132*	<0.01	0.121	0.055	0.006
EGARCH	65 (247)	0.0123**	<0.01	0.095	0.031	0.000
Panel C: Three Calls per Unit of Index						
Unconditional	43 (247)	0.0093	>0.1	0.221	0.203	0.203
90 day	100 (247)	0.0128	>0.1	0.125	0.161	0.177
Adjusted IV	120 (226)	0.0198*	>0.1	0.363	0.363	0.363
EGARCH	65 (247)	0.0185**	>0.1	0.169	0.179	0.185

B. Robustness to the Number of Written Options

The bounds stated in Constantinides and Perrakis (2007) are derived under the restriction that the number of written calls does not exceed the number of units of the index held in the portfolio. In Table VII, we show that with an increase in the ratio to 1.5 options per unit index, the return gains to the *OT* portfolio range from 0.46% to 0.99%, while for a ratio of 2 these gains become 0.62% to 1.32%; in both cases the tests show dominance of *OT* over *IT*. Only when the ratio increases to 3 do the tests fail to establish dominance at conventional significance levels.

C. Robustness to Portfolio Rebalancing

In Section II.C, we described the portfolio rebalancing policy of the index trader and the option trader. The policy depends on the assumed risk aversion

Table VIII
Robustness Tests of Buy-and-Hold Portfolio with 50% in S&P 500

The equally weighted average of all calls violating the upper bound was traded at each date. The symbols * and ** denote a difference in sample means of the *OT* and *IT* traders significant at the 5% and 1% levels in a one-sided bootstrap test with 9,999 trials. Maximal *t*-statistics for the Davidson and Duclos (2000) test are compared to critical values of the Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for nominal levels of 1%, 5%, and 10% with $k = 20$ and $\nu = \infty$. The *p*-values for $H_0 : OT \succ_2 IT$ are greater than 10%, the highest nominal level available in the Stoline and Ury (1979) tables. The *p*-values for the Davidson and Duclos (2006) test are based on 999 bootstrap trials. The *p*-values for $H_0 : IT \not\succeq_2 OT$ are equal to 1.

Volatility Prediction Mode	# Months with Viol. (# Months)	$\hat{\mu}_{OT} - \hat{\mu}_{IT}$ (Annualized)	DD (2000) p -Value $H_0 : IT \succ_2 OT$	DD (2006) p -Value $H_0 : OT \not\succeq_2 IT$ 10% Trimming in Left Tail, Trimming in Right Tail as Below:		
				No Trimming	5% Trimming	10% Trimming
Unconditional	43 (247)	0.0031	<0.01	0.244	0.032	0.000
90 day	100 (247)	0.0048	<0.01	0.175	0.012	0.002
Adjusted IV	120 (226)	0.0065*	<0.01	0.124	0.034	0.000
EGARCH	65 (247)	0.0065**	<0.01	0.086	0.000	0.000

of the trader. For the results presented in Table V, we applied the portfolio rebalancing policy that is optimal for a trader with risk aversion coefficient equal to two. We stress that neither the bounds nor the empirical tests depend on the risk aversion of the trader.

We examine the robustness of our results to the trader's rebalancing policy in three ways. First, we replace the optimal rebalancing policy with a buy-and-hold policy, essentially equivalent to assuming that the transaction cost is infinite. Since the upper and lower stochastic dominance bounds on option prices are independent of the trader's utility, we observe the same number of violations as we do in Table V. The results of the stochastic dominance tests are displayed in Table VIII and are virtually identical to those in Table V.

Second, we set the relative risk aversion coefficient at 10 instead of 2. The change in the risk aversion coefficient only affects the portfolio rebalancing policy. The test results are again virtually identical to those in Table V and are reported in the Internet Appendix. The reason for this robustness is that the no-transactions region is wide and rebalancing is infrequent, irrespective of the degree of risk aversion and irrespective of an exogenously imposed buy-and-hold policy. Third, the test results are virtually unchanged when the starting portfolio is at either boundary of the no-transactions region.

D. Robustness to the Determination of the Bounds

In Table V, the bounds are calculated under the assumption that the futures basis risk is within 0.5% of the index price, consistent with the observation that 95% of all observations over 1990 to 2002 have basis risk within 0.5% of the index price. As a robustness check, we suppress the basis risk and present the results in the Internet Appendix. The bounds become tighter and there

appear to be more violations. Since the call and put upper bounds are lower, the options trader is less selective than before in writing options that violate their upper bounds. As in Table V, the DD (2000) test does not reject the hypothesis $H_0 : OT \succ_2 IT$ and rejects the hypothesis $H_0 : IT \succ_2 OT$. With 10% trimming, the DD (2006) test rejects the hypothesis $H_0 : OT \not\succeq_2 IT$. We conclude that the results in Table V are robust to basis risk.

Next, we set the expected premium on the index at 6% instead of 4%. Since the call and put upper bounds are higher, the options trader is more selective than before in writing options that violate these bounds. We also set the premium at 2%. The results are reported in the Internet Appendix. In all cases, the stochastic dominance results in writing calls are as strong as in Table V. We conclude that the results in Table V are robust to the assumption that the expected premium on the index is 4%.

Finally, we examine whether the unusually high implied volatility after the October 1987 crash affects our results by presenting unusually profitable utility-improving trading opportunities. We exclude from the sample the 7 months from October 1987 to April 1988 and find that the test results are essentially the same as in Table V.

E. Comparison to Alternative Trading Rules

One may downplay our test results in identifying good sell options, arguing that they are a consequence of the widely accepted view that OTM index calls and puts are “expensive” and that even the naïve trading rule of indiscriminately writing all OTM calls and puts is “profitable.” We address this concern by showing that neither a simple heuristic for trading options based on observed option price percentiles for buying low and selling high nor an application of the Bernardo and Ledoit (2000) bounds lead to out-of-sample portfolio returns that stochastically dominate portfolio returns that do not trade in options.

We compare the portfolio return of an index trader with the portfolio return of an option trader who indiscriminately writes all available calls or puts on index futures every month. The results are presented in the first two lines of Table IX, Panel A. Both the DD (2000) and DD (2006) tests are weakly supportive of the hypothesis that the call option trader’s returns stochastically dominate the index trader’s returns. This conclusion is not supported in Panels B and C, where the trader, respectively, writes the top 10% and 2.5% of the calls, because of the reduced sample size.

Next, we compare the portfolio return of an index trader with the portfolio return of a naïve option trader who indiscriminately writes all available puts on index futures every month. The results are presented in the second line of Table IX, Panel A. The DD (2006) test, with any amount of trimming of the right-hand tail, does not reject the hypothesis that the OT portfolio does not dominate the IT portfolio. This suggests that indiscriminate writing of puts does not improve the portfolio return in terms of the stochastic dominance criterion.

Table IX
Returns of Naïve Options Trader and Index Trader

The equally weighted average of all violating options equivalent to one option per share was traded at each date. The symbols *** denotes a difference in sample means of the *OT* and *IT* traders significant at the 1% level in a one-sided bootstrap test with 9,999 trials. Maximal *t*-statistics for the Davidson and Duclos (2000) test are compared to critical values of the Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for nominal levels of 1%, 5%, and 10% with $k = 20$ and $\nu = \infty$. The *p*-values for $H_0 : OT \succ_2 IT$ are greater than 10%, the highest nominal level available in the Stoline and Ury (1979) tables. The *p*-values for the Davidson and Duclos (2006) test are based on 999 bootstrap trials. The *p*-values for $H_0 : IT \not\succeq_2 OT$ are equal to 1.

Trade Type or or Volatility Est. Mode	# Months with Viol. (# Months)	$\hat{\mu}_{OT} - \hat{\mu}_{IT}$ (Annualized)	DD (2000) <i>p</i> -Value $H_0 : IT \succ_2 OT$	DD (2006) <i>p</i> -Value $H_0 : OT \succ_2 IT$ 10% Trimming in Left Tail, Trimming in Right Tail as Below:		
				0%	5%	10%
Panel A: All Quantiles						
Short call	247 (247)	0.0060	<0.01	0.207	0.093	0.007
Short put	247 (247)	0.0078	<0.01	0.280	0.186	0.060
Long call	247 (247)	-0.0403***	<0.01	1	1	1
Long put	247 (247)	-0.0292***	<0.01	1	1	1
Panel B: 10th or 90th Critical Quantile						
Short call	58 (243)	0.0041	<0.01	0.195	0.146	0.023
Short put	67 (243)	0.0034	<0.01	0.331	0.232	0.100
Long call	73 (243)	-0.0149***	>0.1	1	1	1
Long put	95 (243)	-0.0058	>0.1	1	1	1
Panel C: 2.5th or 97.5th Critical Quantile						
Short call	32 (243)	0.0057	<0.01	0.073	0.069	0.094
Short put	36 (243)	0.0022	>0.1	0.359	0.225	0.051
Long call	27 (243)	-0.0038	>0.1	1	1	1
Long put	45 (243)	0.0000	>0.1	0.482	0.477	0.345

The results presented in the last two lines of Table IX, Panels A, B, and C, confirm the obvious, that there is no evidence that indiscriminate buying of calls or puts leads to portfolio returns that stochastically dominate the portfolio return of the index trader.

Overall, the naïve trading rules work well in a quarter of the cases: they work well in writing calls but work poorly in writing puts, buying calls, or buying puts. Thus, the limited success of the naïve trading rules appears to be fortuitous. By contrast, the trading rules based on stochastic dominance bounds work well in practically all cases: they identify good sell calls and puts and find very few good buy calls and puts.

Finally, we compare the portfolio return of an index trader with the portfolio return of an alternative call trader who identifies good buy and sell calls by

applying the Bernardo and Ledoit (BL, 2000) bounds. We apply the BL program specified for the Monte Carlo approach. We simulate the terminal distribution for the S&P futures price under the risk-neutral measure for the future volatility estimates as before. Then, for a given gain-loss ratio we derive the lower or upper bounds on call prices by applying a respective maximization or minimization program to a BL replicating portfolio constrained to satisfy this given gain-loss ratio. Since we apply BL bounds only to calls, neglecting the early exercise feature is innocuous for futures options. Having estimated these bounds, we use them to search for the respective violations of the lower and upper bounds as before. Note that deriving BL bounds for American put options presents technical difficulties beyond the illustrative scope of our application.

The application of the BL bounds requires as input an exogenous parameter, the gain-loss ratio. In Table X, Panels A and C, we report results for a gain-loss ratio of 2. The results for the upper bound are similar to those obtained for the Constantinides and Perrakis (2007) upper bound in Table V; however, the results for the lower bound are disastrous. Even when we increase the gain-loss ratio to 4 (Panels B and D), which by design results in fewer good buys and sells, these results for the lower bound prevail when volatility is estimated unconditionally. At the same time, we observe no evidence for stochastic dominance for the upper bound except for 90-day volatility estimation mode; furthermore, in two cases we observe negative excess returns to OT.

V. Concluding Remarks

We introduce a novel approach for empirical research in option pricing and apply it to S&P 500 index futures options. We search for good sell and good buy American call and put options on S&P 500 index futures by employing stochastic dominance upper and lower bounds on the option prices. We find a substantial number of violations of the call upper bounds, but relatively few violations of the put upper bounds and even fewer of the call and put lower bounds. Since violations of these bounds are too infrequent for statistical inference, we focus on violations of the call upper bounds. We observe that the highest proportion of violations occurs in the region of OTM calls, where the bounds are tight. We also find, however, that the largest number of violations are in the close-to-the-money region and hence likely correspond to liquid options.

We compare the portfolio return of an option trader who writes good sell calls at their bid price with the portfolio return of an index trader who does not trade in the options over the period 1983 to 2006. In out-of-sample tests, our main result is that the return of a call writer stochastically dominates the index trader's return, net of transaction costs and the bid-ask spread. The dominance holds under a variety of methods for estimating the underlying return distribution. It also holds when the trader is allowed to vary her portfolio position by adding other risky assets beyond the index to her portfolio.

Our results are consistent with equilibrium in a segmented market along the following lines. Mutual funds exert price pressure on OTM index puts

Table X
Good Buy and Good Sell Calls Using the Bernardo and Ledoit (2000) Method

The equally weighted average of all violating options equivalent to one option per share was traded at each date. The symbols *, **, and *** denote a difference in sample means of the *OT* and *IT* traders significant at the 10%, 5%, and 1% levels in a one-sided bootstrap test with 9,999 trials. Maximal *t*-statistics for the Davidson and Duclos (2000) test are compared to critical values of the Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for nominal levels of 1%, 5%, and 10% with $k = 20$ and $v = \infty$. The *p*-values for $H_0 : OT \succ_2 IT$ are greater than 10%, the highest nominal level available in the Stoline and Ury (1979) tables. The *p*-values for the Davidson and Duclos (2006) test are based on 999 bootstrap trials. The *p*-values for $H_0 : IT \succ_2 OT$ are equal to 1.

Volatility Prediction Mode	# Months with Viol. (# Months)	$\hat{\mu}_{OT} - \hat{\mu}_{IT}$ (Annualized)	DD (2000) p -Value $H_0 : IT \succ_2 OT$	DD (2006) p -Value $H_0 : OT \not\succeq_2 IT$ 10% Trimming in Left Tail, Trimming in Right Tail as Below:		
				No Trimming	5% Trimming	10% Trimming
Panel A: Upper Bound (Gain to Loss Ratio = 2)						
Unconditional	46 (247)	0.0008	<0.01	0.436	0.150	0.000
90 day	141 (247)	0.0012	<0.01	0.426	0.190	0.048
Adjusted IV	162 (226)	0.0082	<0.01	0.153	0.083	0.013
EGARCH	91 (247)	0.0063	<0.01	0.152	0.056	0.011
Panel B: Upper Bound (Gain to Loss Ratio = 4)						
Unconditional	20 (247)	-0.0012	>0.1	1	1	1
90 day	57 (247)	0.0039	<0.01	0.203	0.102	0.014
Adjusted IV	75 (226)	0.0013	<0.01	0.387	0.325	0.135
EGARCH	27 (247)	-0.0036	>0.1	1	1	1
Panel C: Lower Bound (Gain to Loss Ratio = 2)						
Unconditional	156 (247)	-0.0248***	>0.1	1	1	1
90 day	43 (247)	-0.0080	>0.1	1	1	1
Adjusted IV	11 (226)	-0.0014	>0.1	1	1	1
EGARCH	64 (247)	-0.0123*	>0.1	1	1	1
Panel D: Lower Bound (Gain to Loss Ratio = 4)						
Unconditional	104 (247)	-0.0163**	>0.1	1	1	1
90 day	7 (247)	n/a	n/a	n/a	n/a	n/a
Adjusted IV	2 (226)	n/a	n/a	n/a	n/a	n/a
EGARCH	13 (247)	-0.0015	>0.1	1	1	1

because they buy them as insurance, and overoptimistic speculators exert price pressure on OTM calls because they buy them as a leveraged bet. Furthermore, Garleanu, Pedersen, and Poteshman (2009) argue that dealers inflate the prices of options. As we show in the paper, this presents opportunities for individual investors to write good sell calls and enhance their portfolio returns, net of

transaction costs, in terms of the criterion of stochastic dominance. This can be an equilibrium if the number of such traders and the scale of their trades are sufficiently small that they do not eliminate the overpricing. Large investors such as hedge funds who can potentially eliminate the overpricing do not write these good sell options on a large scale because, as Santa-Clara and Saretto (2009) point out, they face obstacles including margin calls and lack of market depth.

Appendix A: Data

S&P 500 futures have maturities only in months in the March quarterly cycle. Options on the S&P 500 futures have maturities either in a month in the March quarterly cycle (“quarterly options”) or in a month not in the March quarterly cycle (“serial options”). We consider 1-month quarterly options written on 1-month futures and 1-month serial options written on futures with the shortest maturity. We obtain the time-stamped quotes of the 1-month S&P 500 futures options and the underlying 1-month futures for the period February 1983 to July 2006 from the CME tapes.

From futures prices, we calculate the implied S&P 500 index prices by applying the cost of carry relation $F_t = (1 + \gamma)^{-(T^F - t)} R^{T^F - t} S_t + \varepsilon_t$, assuming away basis risk, $\varepsilon_t \equiv 0$. Recall that our goal is to compare the investment policies of the index trader and the option trader. Since both policies stipulate approximately the same stock component, the effects of this component cancel each other out. Also, it is common in empirical work to derive the index value from the index futures; see, for example, Jackwerth and Rubinstein (1996).

We obtain the daily dividend record of the S&P 500 index over the period 1928 to 2006 from the S&P 500 Information Bulletin and convert it to a constant dividend yield for each 30-day period. Before April 1982, dividends are estimated from monthly dividend yields. We obtain the interest rate as the 3-month T-bill rate from the Federal Reserve Statistical Release H.15. We estimate the variance of the basis risk, $\text{var}(\varepsilon_t)$, from the observed futures prices and the intraday time-stamped S&P 500 record obtained from the CME.

We rescale the index price S_t by the multiplicative factor $100,000/S_0$ so that the index price at the beginning of each 30-day period is 100,000. Accordingly, we rescale the futures price, index futures option price, and strike by the same multiplicative factor.

We consider options maturing in 30 calendar days, which results in 247 sampling dates. The 30-day rule eliminates the observation of the October 1987 crash from our sample. Therefore, we use one 40-day period for October 1987 in order to observe the crash. Our results remain unchanged. Since the first maturity of serial options was in August 1987, the first 19 periods occur with quarterly periodicity. Overall, we record 36,921 raw call quotes and 42,881 raw put quotes. After eliminating obvious data errors, we apply the following filters: minimum 15 cents for a bid quote and 25 cents for an ask quote; K/F ratio within 0.96 to 1.08 for calls and within 0.92 to 1.04 for puts; and matching the underlying futures quote within 15 seconds. Part of the data are lost due

to the CME rule of flagging quotes, that is bids (asks) are flagged only if a bid (ask) is higher (lower) than the preceding bid (ask); in addition, no transaction data are flagged. We recover a large part of the data by analyzing the sequence between consecutive bid-ask flags; however, this recovery is not possible in all cases. As a result of the applied filters, we obtain 29,822 quotes for calls and 30,281 quotes for puts in our final sample. These quantities translate into roughly 60 data points for all strikes for either bid or ask prices for an average day.

Appendix B: Trading Policy

We consider calls with moneyness (K/F) within the range 0.96 to 1.08 and puts within the range 0.92 to 1.04. If we observe n call bid prices violating the call upper bound, each with different strike price, then the option trader writes $1/n$ calls of each type with the underlying futures corresponding to the index value of y_0 . The trader transfers the proceeds to the bond account: $x = x_0 + \sum_{i=1}^n C_i/n$ and $y = y_0$.

If we observe n put ask prices violating the put lower bound, each with different strike price, the option trader buys $1/n$ puts of each type and finances the purchase out of the bond account: $x = x_0 - \sum_{i=1}^n P_i/n$ and $y = y_0$.

However, when there is a violation of the put upper bound and the option trader writes puts, the trader also sells one futures contract for each written put. The intuition for this policy may be gleaned from the observation that the combination of a written put and a short futures amounts to a synthetic short call. In fact, the put upper bound in equation (3) is derived from the call upper bound in equation (2) through the observation that if we can write a put at a sufficiently high price we violate the call upper bound by writing a synthetic call.

Finally, when there is a violation of the call lower bound and the option trader buys calls, the trader also sells one futures contract for each purchased call. The intuition is the same as above.

The early exercise policy of a call is based on the function N in equation (2). However, whenever the option trader is short an option, each period we derive the function N based on the *forward-looking* distribution of daily returns, that is, this function is derived under the empirical distribution of the daily index returns between the option trade and the option maturity. Effectively, we endow the counterparty of the option trader with information on the second, third, and fourth moments of the forward distribution, while imposing the first moment. The early exercise policy of a call or put is simplified by the observation that the decision is a function only of time and the ratio of the strike price to the index level.

Appendix C: The Davidson and Duclos (2000, 2006) Tests

The sample counterpart of conditions (4) and (5), applied to the two distributions drawn from their respective populations, is that we must have for every

z in the joint support

$$\hat{D}_{IT}^2(z) - \hat{D}_{OT}^2(z) > 0, \tag{C.1}$$

where

$$\hat{D}_J^2(z) = \frac{1}{N} \sum_{i=1}^N (z - W_{Ji})_+, \tag{C.2}$$

N is the number of paired outcomes, W_{Ji} is the i^{th} outcome of the sample J , and $(x)_+ \equiv \max(x, 0)$. See DD (2000) for further details. Clearly, if (C.1) is violated at any point in the interior of the joint support, the null of nondominance cannot be rejected. On the other hand, by definition (C.1) becomes an equality at one or both endpoints of the support. The DD (2006) test deals with this problem by restricting the set of points over which (C.1) and (C.2) are estimated.

DD (2000) provide a test of the null hypothesis $H_0 : OT \succ_2 IT$ in terms of the maximal and minimal values of the extremal test statistic $\hat{T}(z)$, defined below. The null is not rejected if the maximal value of the statistic is positive and statistically significant *and* the minimal value of the statistic is either positive or negative and statistically insignificant. As opposed to DD (2006), this test may provide evidence for stochastic dominance even if we observe a negative statistic $\hat{T}(z)$.

The variable z denotes the annualized arithmetic return of a trader, where the subscripts IT and OT distinguish between the index trader and the option trader. The statistic $\hat{T}(z)$ is defined as follows:

$$\hat{T}(z) = \frac{\hat{D}_{IT}^2(z) - \hat{D}_{OT}^2(z)}{\sqrt{\hat{V}^2(z)}}, \tag{C.3}$$

where the numerator is given by (C.1) and (C.2), and

$$\hat{V}^2(z) = \hat{V}_{IT}^2(z) + \hat{V}_{OT}^2(z) - 2\hat{V}_{IT,OT}^2(z), \tag{C.4}$$

$$\hat{V}_I^2(z) = \frac{1}{N} \left[\frac{1}{N} \sum_{i=1}^N (z - W_{Ii})_+^2 - \hat{D}_I^2(z)^2 \right], I = IT, OT, \tag{C.5}$$

and

$$\hat{V}_{OT,IT}^2(z) = \frac{1}{N} \left[\frac{1}{N} \sum_{i=1}^N (z - W_{ITi})_+(z - W_{OTi})_+ - \hat{D}_{IT}^2(z)\hat{D}_{OT}^2(z) \right]. \tag{C.6}$$

The maximal and minimal values of the statistic are calculated as a maximum and minimum of (C.3) over a set of points of z , as explained below. Stoline and Ury (1979) provide tables for the nonstandard distribution of the maximal and minimal value of $\hat{T}(z)$ at the 1%, 5%, and 10% levels. In principle, the number of points in this joint support over which the test may be performed

needs to be restricted since a “large” number of these points violate the independence assumption between the $\hat{T}(z)$'s. Therefore, we compute these statistics for 20 points equally spaced in the joint support of W_{IT} and W_{OT} (including the endpoints), which corresponds to $k = 20$ in the Stoline and Ury (1979) tables.

By contrast, DD (2006) develop the concept of *restricted* stochastic dominance in testing the null hypothesis $H_0 : OT \not>_2 IT$. The test derives the *minimal* $\hat{T}(z)$ -statistic over a suitably restricted interval in the joint support for IT and OT . The restriction for the testing interval comes from the observation that a minimal $\hat{T}(z)$ -statistic may not be significant by any distributional standards in the tails of the distribution, be it a sample or a population. It can be easily shown that the leftmost t -statistic is approximately equal to 1, by construction. The numerator of the rightmost t -statistic is simply given by (C.1), the difference of the sample means, which implies that testing for SSD at the largest observed outcome corresponds to testing for the significance in the difference in the sample means; this condition is much stronger than necessary for SSD. Having derived the minimal $\hat{T}(z)$ -statistic in a restricted interval, the DD (2006) test applies a bootstrap procedure to the entire data to derive the p -value for the test as described below.

A necessary condition for applying DD (2006) is that condition (C.1) holds for our sample. By our trading strategies, condition (C.1) holds over the left side of the return distribution. Its validity, therefore, needs to be tested only over the right side, in which case it corresponds to the positivity of the difference of the means of the two samples. We verify this positivity in all cases and, wherever it is satisfied in the sample, we subject it to further verification by block bootstrapping 10 years of results from our data. In almost all cases, the bootstrap results confirm the sign of the means' difference.

The test statistic $\hat{T}(z)$ is the same as in DD (2000) and is given by (C.3) to (C.6). This statistic is computed for the values of z that are sample points within the restricted interval, that is, in this interval we have coupled observations of W_{IT} and W_{OT} , transformed to annualized arithmetic returns. As opposed to the DD (2000) test, there is no restriction on the number of these points and we compute the minimal value of $\hat{T}(z)$ in the restricted interval. It can be shown that $\hat{T}(z)$ is monotonic between the sample points; therefore, the minimal value of $\hat{T}(z)$ may be found only at a sample point. If the minimal value is negative anywhere in the full support, the null of nondominance is accepted with p -value of 1. In this regard, the DD (2006) test is more conservative than the DD (2000) test. The latter test verifies whether an extreme negative $\hat{T}(z)$ is significant for the null $H_0 : IT >_2 OT$ while the former test accepts the null of nondominance. Otherwise, we apply the bootstrap approach for the derivation of the p -values for the null hypothesis, as described in detail in DD (2006) and Davidson (2007). These are simply the number of cases for which the minimal $\hat{T}(z)$ under the bootstrap distribution exceeds the minimum $\hat{T}(z)$ of our sample divided by the number of bootstraps.

In our tests, we use 999 bootstrap replications in order to derive the p -values in the tables. The bootstrap procedure samples all observed coupled values of W_{IT} and W_{OT} under artificial probabilities derived for the empirical likelihood

maximization under the condition that the $\hat{T}(z)$ is set equal to zero at the sample point at which it attains its minimum. See Davidson (2007) for further details.

There is a cost in adopting the DD (2006) null, because, as can be analytically shown, this null cannot be rejected over the entire support of the sample distribution. DD (2006) overcome this problem by restricting the interval over which the null may be rejected to the interior of the support, excluding points at the edges. They then show by simulation that inference on the basis of this restricted interval constitutes the most powerful available inference on the existence of stochastic dominance. In the case of correlated (coupled) samples, the procedure for restricting the interval in the right tail is to start by trimming two pairs of data points: one with the maximal W_{IT} and the corresponding W_{OT} , and one with the maximal W_{OT} and the corresponding W_{IT} . We continue in a similar way until the desired degree of trimming is reached. An analogous procedure is implemented in the left tail. Note that the DD (2006) test results for such a procedure are more conservative than those resulting from trimming pairs of observations in the extremes of the tails of the distribution, irrespective of the sample (*OT* or *IT*) to which these extremes belong.

In the Internet Appendix, using simulated data with characteristics that mirror our sample, we compute the rejection probabilities of the null hypothesis when it is true as well as when it is false. DD (2006) is a weak test without trimming, since it has very low probabilities of rejection of the nondominance null even when it is false. With 5% trimming, the test is still conservative as far as rejecting the false nondominance null. Problems with rejection of the null when it is true occur only for deep OTM options. For this reason, we repeat the stochastic dominance tests for the call upper bound in Panel A of Table V for a restricted moneyness range, that is by removing violating OTM calls outside the range from the sample. The results remain essentially unchanged.

REFERENCES

- Ait-Sahalia, Yacine, and Andrew W. Lo, 2000, Nonparametric risk management and implied risk aversion, *Journal of Econometrics* 94, 9–51.
- Baillie, Richard T., and Tim Bollerslev, 1992, Prediction in dynamic models with time dependent conditional variances, *Journal of Econometrics* 52, 91–113.
- Bernardo, Antonio E., and Olivier Ledoit, 2000, Gain, loss, and asset pricing, *Journal of Political Economy* 108, 144–172.
- Bogle, John C., 2005, The mutual fund industry 60 years later: For better or worse? *Financial Analyst Journal* 61, 15–24.
- Bondarenko, Oleg, 2003, Statistical arbitrage and securities prices, *Review of Financial Studies* 16, 875–919.
- Constantinides, George M., 1986, Capital market equilibrium with transaction costs, *Journal of Political Economy* 94, 842–862.
- Constantinides, George M., Jens C. Jackwerth, and Stylianos Perrakis, 2009, Mispricing of S&P 500 index options, *Review of Financial Studies* 22, 1247–1277.
- Constantinides, George M., and Stylianos Perrakis, 2002, Stochastic dominance bounds on derivatives prices in a multiperiod economy with proportional transaction costs, *Journal of Economic Dynamics and Control* 26, 1323–1352.

- Constantinides, George M., and Stylianos Perrakis, 2007, Stochastic dominance bounds on American option prices in markets with frictions, *Review of Finance* 11, 71–115.
- Coval, Joshua D., and Tyler Shumway, 2001, Expected option returns, *Journal of Finance* 56, 983–1009.
- Davidson, Russell, 2007, Testing for restricted stochastic dominance: Some further results, Working paper, McGill University.
- Davidson, Russell, and Jean-Yves Duclos, 2000, Statistical inference for stochastic dominance and for the measurement of poverty and inequality, *Econometrica* 68, 1435–1464.
- Davidson, Russell, and Jean-Yves Duclos, 2006, Testing for restricted stochastic dominance, Working paper, McGill University.
- Driessen, Joost, and Pascal J. Maenhout, 2007, An empirical portfolio perspective on option pricing anomalies, *Review of Finance* 11, 561–603.
- Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.
- Garleanu, Nicolae, Lasse H. Pedersen, and Allen M. Poteshman, 2009, Demand-based option pricing, *Review of Financial Studies* 22, 4259–4299.
- Jackwerth, Jens C., 2000, Recovering risk aversion from option prices and realized returns, *Review of Financial Studies* 13, 433–451.
- Jackwerth, Jens C., and Mark Rubinstein, 1996, Recovering probability distributions from option prices, *Journal of Finance* 51, 1611–1631.
- Nelson, Daniel M., 1991, Conditional heteroskedasticity in asset returns: A new approach, *Econometrica* 59, 347–370.
- Nolte, Ingmar, 2008, Stochastic dominance tests under test, Working paper, University of Konstanz.
- Oancea, Ioan M., and Stylianos Perrakis, 2009, Jump diffusion option valuation without a representative investor: A stochastic dominance approach, Working paper, Concordia University.
- Perrakis, Stylianos, and Michal Czerwonko, 2006, Transaction costs and stochastic dominance efficiency in the index futures options market, Working paper, Concordia University.
- Rosenberg, Joshua V., and Robert F. Engle, 2002, Empirical pricing kernels, *Journal of Financial Economics* 64, 341–372.
- Rubinstein, Mark, 1994, Implied binomial trees, *Journal of Finance* 49, 771–818.
- Santa-Clara, Pedro, and Alessio Saretto, 2009, Option strategies: Good deals and margin calls, *Journal of Financial Markets* 12, 391–417.
- Stoline, Michael R., and Hans K. Ury, 1979, Tables of the studentized maximum modulus distribution and an application to multiple comparisons among means, *Technometrics* 21, 87–93.
- Tse, Y. K., and Xibin Zhang, 2003, A Monte Carlo investigation of some tests for stochastic dominance, Working paper, Monash University.