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

George M. Constantinides
University of Chicago and NBER

Michal Czerwonko
Concordia University

Jens Carsten Jackwerth
University of Konstanz

Stylianos Perrakis
Concordia University

(JEL G11, G13, G14)

Current draft: November 23, 2008

Keywords: option mispricing; futures options; derivatives pricing; stochastic dominance; transaction costs; market efficiency

We thank Oleg Bondarenko, Wolfgang Buehler, Jim Hodder, Robert Merton, Ingmar Nolte, Myron Scholes, Sorin Sorescu, Suresh Sundaresan, Michael Wolf, 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, the 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 of the University of Chicago; Czerwonko and Perrakis acknowledge financial support from the Social Sciences and Humanities Research Council of Canada. E-mail addresses: gmc@ChicagoGSB.edu, m_czerwo@jmsb.concordia.ca, Jens.Jackwerth@uni-konstanz.de, SPerrakis@jmsb.concordia.ca

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

Abstract

American call and put options on S&P 500 index futures that violate stochastic dominance bounds over 1983-2006 are identified as potentially profitable investments. Call bid prices more frequently violate their upper bound than put bid prices do; evidence of underpriced calls and puts is scant. Inclusion of short positions in overpriced calls, puts, and straddles/strangles in the market portfolio increases the expected utility of any risk averse investor and also increases the Sharpe ratio, net of transaction costs and bid-ask spreads. The results are robust to estimation method of the bounds and relaxation of various assumptions about investors and markets.

A large body of finance literature addresses the mispricing of options. Rubinstein (1994), Jackwerth and Rubinstein (1996), and Jackwerth (2000), among others, observe a steep index smile in the implied volatility of S&P 500 index options that suggests that 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. Constantinides, Jackwerth, and Perrakis (2008) provide empirical evidence that both puts and calls on the S&P 500 index are mispriced by showing that they violate stochastic dominance bounds on European options put forth by Constantinides and Perrakis (2002).

In this paper, we identify American call and put options on the S&P 500 index futures over 1983-2006 that violate the stochastic dominance bounds of Constantinides and Perrakis (2007) as potentially profitable investment opportunities. We then consider the profits that accrue from the exploitation of such mispricing by adopting the appropriate trading policy for a generic investor. In both the identification of mispriced options and the trading policy we recognize realistic trading conditions by incorporating transaction costs, bid-ask spreads, and trading delays (by waiting one quote before entering the position).

We show that trading policies that exploit these violations lead to portfolio returns that *stochastically dominate* (in the second order) portfolio returns that do not exploit these violations. This means that the expected utility of *any* risk averse investor increases when exploiting these violations, independent of the investor's particular endowment and risk-averse utility function. We also show that these policies provide higher Sharpe ratios than policies which do not include options. Below, we highlight novel features of our approach.

First, we use the Chicago Mercantile Exchange (CME) database on S&P 500 futures options, 1983-2006, which is clean and spans a long period. Much of the earlier empirical work on the mispricing of index options is based on data on the S&P 500 index options that comes from two principal sources: the Berkeley Options Database (1986-1995) that provides relatively clean transaction prices, but misses important events over the past 12 years, such as the 1998 liquidity crisis, the dot-com bubble, and its 2000 burst; and the OptionMetrics (1996-2006) data base which, however, is of uneven quality and contains only end-of-day quotes.

Second, we identify mispriced options with a screening mechanism that uses *minimal* assumptions about market equilibrium. This mechanism is based on the stochastic dominance bounds of Constantinides and Perrakis (2007). These bounds identify reservation purchase and reservation write prices such that *any* risk averse investor may increase her expected utility by including the option that violates these bounds in her portfolio. The bounds are valid for any distribution of the underlying asset, including the empirical ones extracted from past data. The bounds also recognize the possibility of early exercise of American options.

A necessary assumption about the market for the validity of these bounds is that there exists a class of traders holding portfolios containing only the S&P 500 index and the risk free asset.[1] Ample evidence exists that this assumption holds for US markets. Numerous surveys have shown that a large number of US investors follow indexing policies in their investments. Bogle (2005) reports that, in 2004, index funds account for about one third of equity fund cash inflows since 2000 and represent about one seventh of equity fund

[1] The mean-variance portfolio theory that gives rise to the Sharpe ratio measure of portfolio performance is based on the assumption that every investor holds the market portfolio and the risk free asset.

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 for several years. We find that any such investor would improve her utility by including in her portfolio an option identified as mispriced by the stochastic dominance bounds.

As a third novel feature, we assess the profitability of our trading policy by employing the powerful statistical tests of stochastic dominance by Davidson and Duclos (2000, 2007) which can deal with option returns even in a setting where we do not make assumptions about the preferences of the investors. These tests compare the profitability of the optimal trading policies of a generic S&P 500 index investor with and without the option, in a setting that recognizes the possibility of early exercise of the futures option. The comparisons are valid from the perspective of any risk averse investor. By contrast, the ubiquitous Sharpe ratio measure of portfolio performance (which we present as well) is valid only from the perspective of a mean-variance investor and suffers from well known problems when used to assess non-normal returns such as those encountered in portfolios that include options.

Finally, both the bounds employed in detecting mispriced options and the optimal trading policies with and without the option explicitly take into consideration bid-ask spreads, trading costs and trading delays. Once a trading opportunity is detected, we execute the trade by buying at the next ask price or selling at the next bid price.

Our tests are non-parametric, in the sense that we do not assume any particular distribution for the underlying asset returns. We use historical data on the underlying S&P 500 index returns in order 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 one of these estimates, we evaluate the corresponding bounds over the period 1983-2006 and then identify the observed S&P 500 futures options prices that violate them. For each violation, we identify the optimal trading policy of a generic investor with and without the mispriced option, using the observed path of the underlying asset till 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 of violations.

We find a substantial number of violations of the upper bounds, but relatively few violations of the lower bounds. Since the frequency of violations of the lower bounds is too low for statistical inference, we focus on violations of the upper bounds. In all our tests, we find that the portfolio that includes the mispriced options dominates the portfolio without the options. The dominance holds for all empirical estimates of the underlying asset distribution and is robust to the relaxation of several assumptions about the investor and the market. The results are thus strongly supportive of mispricing. The results also demonstrate the ability of the stochastic dominance bounds to identify mispriced options, in contrast to the failure of a naïve heuristic which we based on observed option price percentiles.

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. In Section III, we present the empirical results and discuss their robustness. We conclude in Section IV.

I. Restrictions on Futures Option Prices Imposed by Stochastic Dominance

We assume that market agents are heterogeneous and 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 only in two of them, a bond and a stock with the natural interpretation as a market index.[2] 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, $(1+\gamma)\left(S_{t+1}/S_t\right)$, 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 cash-settled 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 $F_t = \left(1+\gamma\right)^{-\left(T^F - t\right)} R^{T^F - t} S_t + \varepsilon_t, \quad t \leq T^F, \quad \left|\varepsilon_t\right| \leq \bar{\varepsilon}$, 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

[2] Essentially, we model buy-and-hold investors who trade infrequently and incur low transaction costs. At least for large investors who earn a fair return on their margin, transaction costs are even lower in the index futures market than the stock market. In practice, however, buy-and-hold investors invest in the stock and bond markets because of the inconvenience, cost, and basis risk of the frequent rolling over of short-term futures contracts and the illiquidity of long-term futures and forward contracts.

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. 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.

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 this paper.

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 $(t = 0, 1... , T')$, 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'.[3] 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, IT, and denote her maximized expected utility by $V_0^{IT}\left(x_0, y_0\right)$.

In the second scenario, as in the first 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.[4,5] We assume that the trader may not trade the call option thereafter.[6] At each trading date t ($t = 0,\ 1...\ ,T$) the trader is informed whether or not 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 $x_t - \left(F_t - K\right)$. The trader makes sequential investment decisions with the objective to

[3] Alternatively, the objective may be the maximization of the discounted sum of the utility of consumption $u_t(c_t)$ at each trading date, including the terminal date. In this case, the terminal date may be finite or infinite. Although the Constantinides and Perrakis (2007) bounds are derived under the terminal wealth objective, they remain valid without any reformulation under the alternative objective.

[4] 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.

[5] 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.

[6] The reservation write price of a call is derived under this constrained policy. Define $\overline{C}$ as an upper bound on the reservation write price of a call. Under this constrained policy, the investor increases her expected utility by writing a call at price $\overline{C}$ and refraining from trading the call thereafter. If the constraint on trading the call is relaxed, the policy which the investor follows under the constraint policy remains feasible and increases her expected utility by writing a call at price $\overline{C}$. Therefore, $\overline{C}$ remains an upper bound on the reservation write price of a call. Whereas the upper bound may be tightened when the constraint on trading the call is relaxed, there is no known tighter bound that is preference free. For further discussion on these bounds, see Constantinides and Perrakis (2007).

maximize her expected utility, $E[u_{T'}(W_{T'})]$. We refer to this trader as the *option* (plus index and bond) trader, OT, and denote her maximized expected utility by $V_0^{OT}\left(x_0 + C, y_0\right)$.

For a given pair $\left(x_0, y_0\right)$, we define the reservation write price of a call as the value of C such that $V_0^{OT}\left(x_0 + C, y_0\right) = V_0^{IT}\left(x_0, y_0\right)$. The interpretation of C is the write price of the call at which the trader with initial endowment $\left(x_0, y_0\right)$ is indifferent between writing the call or not. Constantinides and Perrakis (2007) state a tight upper bound on the reservation write price of a 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:

$$\overline{C}\left(F_t, S_t, t\right) = \frac{1+k_1}{1-k_2}\max\left[N\left(S_t, t\right),\ F_t - K\right], \qquad t \le T. \tag{1}$$

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

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

The economic interpretation of the call upper bound is as follows. If we observe a call bid price above the reservation write price, $\overline{C}$, 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. Specifically, without transaction costs on the index, the upper bound is $\max\left[N\left(S_t, t\right),\ F_t - K\right]$; with transaction costs on the index, the upper bound merely increases by the multiplicative factor $\left(1+k_1\right)/\left(1-k_2\right)$. 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 which is optimal for the index trader but is generally suboptimal for the option trader. This policy incurs very low transaction costs because the trader trades infrequently, as shown in Constantinides (1986).

If we further assume that the trader can buy a call at price $\overline{C}\left(F_t, S_t, t\right)$ or less and trade the futures and do so costlessly, we obtain the following put upper bound:[7]

$$\overline{P}\left(F_t, S_t, t\right) = \overline{C}\left(F_t, S_t, t\right) - R^{-(T-t)}F_t + K, \quad t \le T. \tag{3}$$

The interpretation of the put upper bound is as follows. If we observe a put bid price above the reservation write price $\overline{P}$, then any trader can increase her expected utility by writing the put.

Constantinides and Perrakis (2007) also state a lower bound on the reservation purchase price of an American futures put option:

[7] We prove equation (3) by noting that an investor achieves an arbitrage profit by buying a call at $\overline{C}\left(F_t, S_t, t\right)$, writing a put at $P, P > \overline{P}\left(F_t, S_t, t\right)$, selling one futures, 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.

$$\underline{P}(F_t, S_t, t) \equiv \max\left[K - F_t,\ \frac{1-k}{1+k} M(S_t,\ t)\right], \quad t \le T\ . \tag{4}$$

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

$$\begin{aligned} M(S,t) &= (R_S)^{-1} E\left[\max\left[K - (1+\gamma)^{-(T^F - t - 1)} R^{T^F - t - 1} S_{t+1} - \bar{\varepsilon},\ M(S_{t+1}, t+1)\right] \mid S_t = S\right], t \le T - 1 \\ &= 0, \quad t = T. \end{aligned} \tag{5}$$

If we further assume that the trader can write a put at price $\underline{P}(F_t, S_t, t)$ or more, and trade the futures and do so costlessly, then we obtain the following call lower bound:[8]

$$\underline{C}(F_t, S_t, t) = \underline{P}(F_t, S_t, t) + R^{-(T-t)} F_t - K, \quad t \le T\ . \tag{6}$$

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

[8] We prove equation (6) by noting that an investor achieves an arbitrage profit by writing a put at $\underline{P}(F_t, S_t, t)$, buying a call at $C, C < \underline{C}(F_t, S_t, t)$, selling one futures, and lending $K - R^{-(T-t)} F_t$.

IT and OT portfolios in terms of their Sharpe ratios and in terms of the criterion of second order stochastic dominance.

II.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 February 1983-July 2006 from the Chicago Mercantile Exchange (CME) tapes. This results in 247 sampling dates. We obtain the interest rate as the three-month T-bill rate from the Federal Reserve Statistical Release. The data sources are described in further detail in Appendix A.

For the daily index return distribution, we use the historical sample of log 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 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.[9] We estimate the conditional volatility in three different ways: (1) the sample standard deviation over the preceding 90 trading days;[10] (2) the ATM

[9] We also estimated the unconditional volatility over the 24 years prior to January 1983. The results remain essentially unchanged and are not reported.

[10] We have also estimated the conditional volatility over the preceding 360 days. The results remain essentially unchanged and are not reported.

implied volatility (IV) on the preceding day, adjusted by the mean prediction error for all dates preceding the given date (typically some 3%);[11] and (3) 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 till the option expiry date.[12] We estimate the 3rd and 4th 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.

[Table I about here]

II.B Calibration of the index returns tree and calculation of the option bounds

We model the path of the daily index return till the option expiration on a T-step tree, where T is the number of trading days in that particular month.[13] The tree is recombining with m branches emanating from each node. Every month we calibrate the tree by choosing the number of branches, spacing, and transition probabilities at each node to match the first four moments of the daily index return distribution, as described in Appendix B.

---

[11] 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.

[12] 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 pre-sample 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.

[13] For example, if the 3rd Friday of July is on July 27, we record the price of the July option on June 27, which is 30 calendar days earlier. (If June 27 is a holiday, we record the price on June 26.) If there are 21 trading days between June 27 and July 27, we model the path of the daily index return till the option expiration on a 21-step tree.

The upper bounds on the call and put prices are given in equations (1)-(3) and the lower bounds are given in equations (4)-(6). We numerically calculate the bounds by iterating backwards on the calibrated tree.

II.C Portfolio construction and trading

For each monthly stock return path, we employ the following trading 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, as derived in Constantinides (1986) and extended in Perrakis and Czerwonko (2006) to allow for 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 boundaries for the following parameter values: one-way transaction cost rate on the index of 0.5%; annual return volatility of the index of 0.1856, the sample volatility over 1928-1983; interest rate equal to the observed 3-month T-bill date; risk premium 4%; and constant relative risk aversion coefficient of 2.[14] 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.

[14] We clarify that the upper and lower stochastic dominance bounds on option prices apply to any risk averse trader, independent of her particular form of utility function. Only the trading policy is affected by the choice of risk-aversion, but only slightly, because the no transactions region is wide and trading is infrequent. In our empirical work, we present results for a trader with constant relative risk aversion coefficient 2 and 10.

We repeated our tests by replacing the optimal trading policy with a buy-and-hold policy. The results are virtually identical and are not reported here.

For the option trader (who manages a portfolio of the option, 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.[15] However, this portfolio composition changes, depending on the assumed position in futures options, as explained in Appendix C. We employ the trading policy which 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 where the basis risk bound, $\bar{\varepsilon}$, is 0.5% of the index price. Over the years 1990-2002, 95% of all observations have basis risk less than 0.5% of the index price. For reference purposes, we also consider the case $\bar{\varepsilon} = 0$. As expected, when we suppress the basis risk, the bounds are tighter and there appear to be many more violations.

II.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, respectively. Our goal is to test whether the portfolio profitability of the index and option traders are statistically different in the months in which we observe violations of the bounds.

[15] 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.

In our first set of tests, we compare the Sharpe ratios of the two portfolios. Despite the well-known limitations of the Sharpe ratio, we report these results because the Sharpe ratio is a popular measure of portfolio performance. We use the approach of Jobson and Korkie (1981) with the Memmel (2003) correction that accounts for different variances of the two portfolios. Details of the test are described in Appendix D.

In our second set of tests, 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:

$$D_{IT}^{2}(z) - D_{OT}^{2}(z) \geq 0, \tag{7}$$

where

$$D_{J}^{2}(z) = \int_{\underline{z}}^{z} (z - x)\, dF_{J}(x)\,, \tag{8}$$

$J = OT, IT$, $F_J(x)$ is the cumulative distribution function, 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 that 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 that 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 D.[16] 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 $H_0 : OT \not\succ_2 IT$ which states 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 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. Likewise, we test the converse null hypothesis $H_0 : IT \not\succ_2 OT$ 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 (2007) (DD (2007)), using the algorithm developed by Davidson (2007). The test is described in Appendix D. Again, the test requires that returns be serially uncorrelated, an assumption that holds well in all our return series.

The power of the DD (2007) test is low, unless one trims the tails of the paired outcomes. Therefore, we trim 10% of the paired outcomes in the left tail of our sample distributions which affects both IT and OT similarly and is therefore innocuous. Restricting the range further by trimming outcomes on 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 towards rejection of the null, because IT tends to produce superior results to OT when the return of the underlying asset is high. For this

[16] We choose this test because, unlike several alternatives, it applies to correlated samples and is more powerful than other well-known tests. See Tse and Zhang (2003).

reason, we present in our tables results with no trimming on the right and also report the trimming on the right that is necessary to reject the null hypothesis of no dominance at the 5% and 1% significance levels, respectively, whenever a rejection at these levels may be attained. To facilitate interpretation, we perform all our statistical tests on annualized arithmetic returns on the wealth of OT and IT investors.[17]

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 overpriced calls, puts, or straddles/strangles at their bid price with the portfolio return of an index trader who does not trade in the options over the period 1983-2006. We find that the Sharpe ratio of the option trader's return is higher than the Sharpe ratio of the index trader's return and the difference is statistically significant in most cases. We also 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. In Section III.C, we establish that the empirical results are robust. In Section III.D, we demonstrate that trading policies triggered by violations of the stochastic dominance bounds consistently outperform the naïve filter

[17] 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%.

rule of buying low and selling high. In Section III.E, we discuss the robustness of the tests to time-variation of the bounds.

III.A A first look at the results

In Figure 1, we plot the four bounds for one-month options, expressed in terms of the implied volatility, as a function of the moneyness, K/F. We set $\sigma = 20\%$ and $\overline{\varepsilon} = 0$. 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.

[Figures 1 and 2 about here]

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-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.

[Table II about here]

In Table II, the violations are shown as a proportion of the quotes in each moneyness range. It is clear that, for all methods of estimating the bounds, there is a large proportion of violations. For the moneyness range $1.03-1.08$, a large proportion of the available quotes violate the corresponding bound for all estimation methods.

[Table III about here]

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 1.01-1.03 moneyness range and not in the 1.03-1.08 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, 0.99-1.03.[18] Our data also shows that the average size of the mispricing is between 1% and 2% of the upper bound for most methods of estimating the bounds. In the stochastic dominance tests, the power of the tests depends, by construction, on the proportion of months with observed violations.

III.B Differences in Sharpe ratios and stochastic dominance

We apply our statistical tests to all the 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.[19] In Table IV, Panels A and B, we present the cases of call and

[18] The exchange regulations specify that the minimum number of available contracts must be at least 20 for each quote.
[19] We find significant differences in the Sharpe ratios. Our results would have been even stronger, if we had applied our statistical test only to the months that the OT trader trades in options.

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 upper call bound than of the upper put bound because the upper call bound is tighter than the upper put bound, as we observed in Figure 1.

[Table IV about here]

The annual Sharpe ratio of the call trader's return is at least 6.8% higher than the Sharpe ratio of the index trader's return, irrespective of the mode of predicting the volatility, as an input to the call upper bound. When the call upper bound is calculated using the adjusted IV or the EGARCH volatility, the difference in Sharpe ratios exceeds 9% annually and is statistically significant at the 1% level. There are far fewer violations of the put upper bound and, therefore, the results are statistically weaker. Nevertheless, when using the unconditional prediction of volatility as an input to the put upper bound, we find 23 violations of the put upper bound and the put trader's portfolio has a Sharpe ratio that exceeds the index trader's portfolio by 2.6%, which is statistically significant at the 1% level. These results on Sharpe ratios motivate and reinforce our main results on stochastic dominance which are discussed next.

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 upper call bound and the put upper bound. The results are not reported in the table.

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 IV, Panel A, the p-values are lower than 1% and the hypothesis is rejected when the option trader writes overpriced calls. In Panel B, the first two rows, 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 $H_0 : IT \nsucc_2 OT$ which states that either the option trader's return dominates the index trader's return or that neither return dominates the other. We apply the DD (2007) test and obtain p-values of one for both the upper call bound and the put upper bound. The results are not reported in the table.

Finally, we consider the hypothesis $H_0 : OT \nsucc_2 IT$, which states that either the index trader's return dominates the option trader's return or that neither return dominates the other. As we explained earlier, the power of the DD (2007) 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 any trimming on the right tail, the test has low power and the p-values are high in the two panels of Table IV. Without trimming, we reject at the 10% level for one case in Panel A. We also report the trimming on the right that is necessary to reject the null hypothesis of no dominance at the 5% and 1% significance levels, respectively, while maintaining the 10% trimming on the left. In Panel A, a moderate amount of trimming suffices to reject the null. The results are inconclusive in Panel B, largely because there are few months in which we observe violations of the put upper bound.

Overall, the results in Table IV imply that the relatively large number of violations of the call upper bound by call bid prices leads to a trading policy that generates a higher Sharpe ratio than the passive index trader's portfolio; and the option trader's return stochastically dominates the index trader's return. The results point in the same direction for violations of the put upper bound by put bid prices, but the statistical significance is weak because there are very few months in which we observe violations of the put upper bound.

[Table V about here]

Next, we explore the performance of the policy of writing overpriced calls through the policy of writing straddles and strangles. Straddles/strangles are popular trading policies and have been previously investigated in the literature. For example, Coval and Shumway (2001) show that a long ATM straddle on the S&P 500 index or the S&P 100 index produces substantial negative returns.[20] Each month, we look for call bid prices that lie above the upper call bound. If we find at least one call bid price that lies above the upper call bound *and* if we find at least one put bid price with same moneyness (irrespective of whether the put bid price violates the put upper bound or not) we proceed as follows. We short equal fractions of the calls that violate the call upper bound, such that the fractions add up to one; we short equal fractions of the puts for which we have bid prices, such that the fractions add up to one; and we sell one futures contract on the index. The results are reported in Table V, Panel A. The annualized Sharpe ratio differentials are large and significant at the 1% level. These results are consistent with the results of Coval and

[20] The paper of Coval and Shumway (2001) focuses on the relation between the CAPM beta and the return of straddles/strangles and, as such, differs from our goal of measuring the performance of straddles/strangles, net of bid-ask spreads and through the broader criterion of stochastic dominance.

Shumway (2001). The DD (2000) test does not reject the hypothesis $H_0 : OT \succ_2 IT$. This result is not reported here. The test of the hypothesis $H_0 : IT \succ_2 OT$ is inconclusive. Finally, the DD (2007) test (with trimming) rejects the hypothesis $H_0 : OT \not\succ_2 IT$. We conclude that the results in Table V, Panel A, are largely consistent with those in Table IV.

In Table V, Panel A, we note that the number of months in which straddles are traded is significantly lower than the number of months in which calls are sold in Table IV, Panel A. We increase the number of months in which straddles are traded by relaxing the requirement that the put sale has to occur at the same strike price as that of the call that triggers the violation. Instead, we require that the moneyness of the put remains within 0.98-1.02 times the moneyness of the triggering call. The first put quote within this bound following the call violations is included in the ensuing strangle position. The results for this approach are presented in Table V, Panel B. Compared to Table V, Panel A, we observe a systematic increase in the Sharpe ratios and improvement in the stochastic dominance test results.

III.C Robustness tests

In Tables VI-VIII, we demonstrate that the results of Table IV are robust. Table VI differs from Table IV only in that the basis risk is set at zero, $\overline{\varepsilon} = 0$, instead of bounding the basis risk by $\overline{\varepsilon} = 0.5$. There are now more options across the board violating the bounds because all the bounds become tighter: the upper bounds are lowered and the lower bounds are raised. We present the cases of call and put bid prices violating their upper bound. We do not present results for the cases when the call and put ask prices violate their lower bound because we still do not have a sufficient number of such violations to be able to make statistical inference.

[Table VI about here]

Since the upper call and put bounds are lower, the options trader is less selective than before in writing options that violate their upper bounds. The DD (2000) test does not reject the hypothesis $H_0 : OT \succ_2 IT$ and rejects the hypothesis $H_0 : IT \succ_2 OT$. Finally, the DD (2007) test (with trimming) rejects the hypothesis $H_0 : OT \nsucc_2 IT$. We conclude that the results in Table VI are consistent with those in Table IV.

[Table VII about here]

Table VII differs from Table IV only in that the relative risk aversion coefficient is set at 10 instead of 2. 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 in Table VII as we do in Table IV. The change in the risk aversion coefficient does change the boundaries of the no-transactions region and, therefore, the trading policy of the index trader and the option trader. The results of the stochastic dominance tests in Table VII are virtually identical to those in Table IV, thus confirming our earlier conjecture that these test results are unaffected by the choice of investor risk aversion parameter. As further confirmation, we repeated the stochastic dominance tests when the trader follows a buy-and-hold policy on the index. The results are virtually identical as in Table IV and are not reported here.

[Table VIII about here]

Table VIII differs from Table IV only in that the expected premium on the index is set at 6% instead of 4%. Since the upper call and put bounds are higher, the options trader is more selective than before in writing options that violate their upper bounds. The differences in Sharpe ratios are comparable to those in Table IV and always significant at the 1% level for writing calls. The stochastic dominance results in writing calls are as strong as in Table IV. We conclude that the results in Table IV are robust to the assumption that the expected premium on the index is 4%.

Next, we exclude from the sample the seven months from October 1987 to April 1988 in order to abstract from the crash since unusually high post-crash implied volatility might have presented unusually profitable trading opportunities. Since both the stochastic dominance results and Sharpe ratios are essentially the same as in Table IV, we do not report separately these results.

We repeated the tests in Table V on straddles and strangles but with the basis risk set at zero, $\bar{\varepsilon} = 0$, instead of bounding the basis risk by $\bar{\varepsilon} = 0.5$. The results are consistent with those in Table V. We also repeated the tests in Table V with the expected premium on the index set at 6% instead of 4%. Again, the results are consistent with those in Table V.

III.D Comparison with naïve trading policies

In the introduction, we cited evidence that index calls and puts are generally overpriced. This motivates comparison of the portfolio return of an index trader with the portfolio return of a naïve option trader who indiscriminately writes all available calls on index futures every month. The results are presented in the first line of Table IX, Panel A. The OT portfolio has significantly higher Sharpe ratio than the IT portfolio. However, the DD (2007) test requires significant right-hand trimming to reject the hypothesis that the OT portfolio does not dominate the IT portfolio. By contrast, in Tables IV and V, the DD (2007) test requires lesser right-hand trimming to reject the hypothesis that the OT portfolio does not dominate the IT portfolio. This suggests that even indiscriminate writing of calls improves the portfolio return in terms of the stochastic dominance criterion. However, the evidence is stronger in favor of judicious writing of calls that violate the upper bounds.

[Table IX about here]

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 OT portfolio has significantly higher Sharpe ratio than the IT portfolio. However, the DD (2007) 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.

The results presented in the last two lines of Table IX, Panel A, confirm the obvious. Indiscriminate buying (instead of writing as above) of calls or puts leads to portfolio returns with lower Sharpe ratio. Furthermore, there is no evidence that such portfolio returns stochastically dominate the portfolio return of the IT.

Finally, we explore tighter naïve trading policies, motivated by the observation that the option trading policies triggered by violations of the stochastic dominance bounds superficially resemble the trading policy of buying (or selling) an option when its implied volatility (IV) is at the low (or, high) end of the IV distribution of options within a certain range of moneyness. We demonstrate that a particular form of this naïve trading policy consistently underperforms the earlier trading policies triggered by violations of the stochastic dominance bounds.

We derive the 90th, 97.5th, 10th, and 2.5th percentiles of the IV distribution for a given range of moneyness by applying the quantile regression method of Yu and Jones (1998).[21] Table IV, Panels B and C, present the results for the naïve trading policy. In all trading policies, we mirror the policies applied for the option trader (OT). We observe that the number of cross-sections for which we find 'quantile violations' is relatively low. This observation is caused by the clustering of violations in some cross-sections. We conclude that the naïve trading policy detects parallel shifts in the implied volatility instead of singling out unusual observations in the majority of cross-sections. These shifts appear to be inefficient: the naïve bounds only capture some of these parallel shifts in implied volatility, namely violations at the top. The naïve trading policy performs well on the sell side but performs disastrously on the buy side, as shown by the stochastic dominance statistics and Sharpe ratios. The stochastic dominance results for writing calls in Table IV consistently

---

[21] The quantile regression is a kernel regression in two dimensions, in our case in the dimensions of moneyness and IV. As is usual in a kernel regression, the critical part is in determining the kernel bandwidth. To determine this quantity in the moneyness dimension, we use the Leave-One-Out method, as described in Härdle (1990), for which we use the transformation given in Table 1 in Yu and Jones (1998). To determine the bandwidth in the implied volatility dimension, we use (12) in Yu and Jones (1998). As our sample to derive the critical quantile function, we use five past observations with 30 days to maturity. Using ten past observations yields similar results, not reported here. We verified that in-sample the likelihood of observations outside any critical quantile q is close to $\min\left(q,\left|1-q\right|\right)$.

outperform the corresponding results for the naïve filter rule in Table IX, while the results for writing puts in Table V are better or at par.

III.E Robustness to the variation of the bounds

In our empirical work, the parameters and, therefore, the bounds are re-estimated every month. Since the stochastic dominance tests do not recognize this variation of the bounds, we bootstrapped every monthly distribution of the OT and IT portfolio returns from the daily returns of their respective portfolios to option expiration. A test of the hypothesis $H_0 : OT \succ_2 IT$ against the alternative $H_A : OT \not\succ_2 IT$ takes the form of comparing the two bootstrapped distributions at each time point and verifying with a binomial sign test whether the number of points at which $OT \succ_2 IT$ is significantly above 50% of the entire sample. This turns out to be the case for all four methods of estimating the call and put upper bounds. We conclude that our results are robust to this variation of the bounds.

IV Concluding Remarks

We introduce a new approach for empirical research in option pricing and apply it to S&P 500 index futures options. We search for mispriced American call and put options on the S&P 500 index futures by employing stochastic dominance upper and lower bounds on the prices of options. We estimate the bounds with four different methods, all of them based on observable measures at the time option trading takes place. We identify call and put bid prices on index futures that violate the upper bounds and call and put ask prices that violate the lower bounds. We find a substantial number of violations of the upper bounds, but relatively few violations of the lower bounds. Since the frequency of violations of the lower bounds is too low for statistical inference, we focus on violations of the upper bounds.

We observe that the highest proportion of violations occur 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, liable to correspond to more liquid options.

We compare the portfolio return of an option trader who writes overpriced 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 1983-2006. In out-of-sample tests, our main result is that the return of a call or put writer stochastically dominates (in second order) the index trader's return, net of transaction costs and the bid-ask spread. The dominance holds under a variety of methods in estimating the underlying return distribution.

We also find that the Sharpe ratio of the call trader's return is uniformly higher than the Sharpe ratio of the index trader's return and is uniformly statistically significant. Finally, the policy of writing straddles/strangles produces returns that strongly stochastically dominate the index trader's return and have substantially higher Sharpe ratios. The results are supportive of the hypothesis that the options identified by violations of the stochastic dominance bounds are mispriced.

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 one-month quarterly options written on one-month futures and one-month serial options written on futures with the shortest maturity. We obtain the time-stamped quotes of the one-month S&P 500 futures options and the underlying one-month futures for the period February 1983-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, $\varepsilon_t \equiv 0$.[22] We obtain the daily dividend record of the S&P 500 index over the period 1928-2006 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 three-month T-bill rate from the Federal Reserve Statistical Release. We estimate the variance of the basis risk, $\mathrm{var}(\varepsilon_t)$, 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.

[22] 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 a common empirical approach to derive the index value from the index futures; see, for example, Jackwerth and Rubinstein (1996).

We consider options maturing in 30 calendar days, which results in 247 sampling dates.[23] 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 0.96-1.08 for calls and within 0.92-1.04 for puts; and matching the underlying futures quote within 15 seconds. Part of the data is lost due to the CME rule of flagging quotes, i.e. bids (asks) are flagged only if a bid (ask) is higher (lower) than the preceding bid (ask); in addition, no transaction data is 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: Calibration of the index return tree

For every month, we model the paths of the daily index return on a recombining tree with m branches emanating from each node. The objective is to match as closely as possible the first four moments of the daily return distribution. As explained in Section II.A, we fix the mean and use the estimated volatility from one of our four methods. We use as the third and fourth moment the observed sample moments over the 90 preceding calendar days.

In the first step of our algorithm, we pick an odd value for the number of branches m and group the sample of daily returns in a histogram with m bins of equal length (on the log

[23] 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.

scale) such that the extreme bins are centered on the extreme observed returns. The center of each bin then becomes a state in the lattice, with the ordered states and the corresponding probabilities denoted respectively as x_i and p_i, $i = 1...m$. Note that this equidistant log-scale and an odd value for the number of branches m are necessary for the lattice to recombine.[24]

In a second step, we match our moments by fixing the number of branches m and matching the first three moments by changing the spacing (via parameters a and b) and the probabilities via parameter c. The forth moment is then matched by changing the number of branches, m.

We derive the required parameters a, b, and c by solving the following set of three non-linear equations that are simply three moment conditions for the constants a, $b,$ and c:

$$\sum_{i=1}^{m} p_i^* \exp\left(ax_i + b\right) - \exp\left(\hat{\mu}\right) = 0$$

$$\sum_{i=1}^{m} p_i^* \left[\exp\left(ax_i + b\right)\right]^2 - \exp\left(\hat{\mu}\right)^2 - \hat{\sigma}^2 = 0 \tag{B.1}$$

$$\sum_{i=1}^{m} p_i^* \left[\exp\left(ax_i + b\right) - \exp\left(\hat{\mu}\right)\right]^3 - \hat{\mu}_3 \hat{\sigma}^3 = 0$$

where $\exp\left(\hat{\mu}\right)$ and $\hat{\sigma}^2$ are the first and second target moments, respectively; $\hat{\mu}_3$ is the sample skewness; and $p_i^* \equiv \dfrac{p_i + c\underline{1}_{\left(i \geq n^*\right)}\underline{1}_{\left(p_i \neq 0\right)}}{\sum_{i=1}^{m}\left(p_i + c\underline{1}_{\left(i \geq n^*\right)}\underline{1}_{\left(p_i \neq 0\right)}\right)}$, where $\underline{1}_{(.)}$ is the indicator function, n^* is the index to this x_i which brackets from above the target expected log-return $\hat{\mu}$. The first indicator function ensures that the constant c is added only to the probabilities in the

[24] We did not build our lattice by discretizing a kernel-smoothed distribution because this method requires a substantially larger lattice. We did not adopt the Edgeworth/Gram-Charlier binomial lattice methodology, as in Rubinstein (1998), because it sometimes results in negative probabilities.

right tail of the distribution; the second one ensures that the constant c is added only to the positive probabilities. Note that the affine transformation of the log-states x_i preserves the equal distance between the adjacent states. The constant a ensures the desired scale of the log-states x_i, the constant b ensures the desired location of these states, while the constant c increases or decreases the probabilities in the right tail relative to the left one to match the desired skewness.[25]

To match the fourth sample moment $\hat{\mu}_4$, we search over m, the number of nodes in the lattice. With each new m the initial distribution derived from a histogram changes, providing some variability in the fourth moment after the adjustments resulting from solving (B.1). After a search over a range of m's, we pick this distribution which has the lowest absolute difference between its kurtosis and the sample kurtosis $\hat{\mu}_4$. This search procedure results in very small errors in matching $\hat{\mu}_4$ for the data that we use while we obtain the exact match in the first three moments. For the four volatility prediction modes which we apply in our work, the relative error on the fourth moment had the following characteristics: median 0.003%, 99$^{\text{th}}$ percentile 0.105%, maximum 1.659% across 973 observations while we constrained the lattice size m to be no larger than 201.[26]

Appendix C: Trading policy

We consider calls with moneyness (*K/F*) within the range 0.96-1.08 and puts within the range 0.92-1.04. If we observe n call bid prices violating the call upper bound, each with

[25] Note that the presented adjustment of the probabilities in the right tail may not yield an admissible solution, i.e. we may end up with some negative probabilities. If this is the case, we introduce an analogous adjustment in the left tail of the distribution.

[26] This lattice size appears unattractive to derive recursive conditional expectations. However, the use of fast Fourier transforms results in a fairly short processing time. See Cerny (2004).

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 upper put 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 upper put bound in equation (3) is derived from the upper call bound in equation (2) through the observation that if we can write a put at a sufficiently high price we violate the upper call bound by writing a synthetic call.[27]

Finally, when there is a violation of the lower call 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). The early exercise policy of a put is based on the function *M* in equation (5).

However, whenever the option trader is short an option, each period we derive the functions *N* and *M* based on the *forward-looking* distribution of daily returns, i.e. these

[27] In implementing the trading policy of either writing puts or buying calls, the option trader buys or sells a futures contract as well and this violates the assumption made in Section I that the option trader does not trade in futures. Even when we relax the assumption on trading in futures, in practice, traders manage their portfolio by trading in the index because of the inconvenience and cost of the frequent rolling over of short-term futures contracts and the illiquidity of long-term futures and forward contracts.

functions are 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 2nd, 3rd, and 4th 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 D: The Sharpe ratio and the Davidson-Duclos (2000, 2007) tests

For the Sharpe ratio tests, we use the approach of Jobson and Korkie (1981) with the Memmel (2003) correction that accounts for different variances of the two portfolios. Specifically, given the sample of N realizations of the index trader's (IT) and option trader's (OT) portfolio outcomes with $\hat{\mu}_{OT}$, $\hat{\mu}_{IT}$, $\hat{\sigma}^2_{OT}$, $\hat{\sigma}^2_{IT}$, $\hat{\sigma}_{IT,OT}$ as their estimated excess means, variances, and covariances, we test the hypothesis $H_0 : \hat{\mu}_{OT}\hat{\sigma}_{IT} - \hat{\mu}_{IT}\hat{\sigma}_{OT} \leq 0$ with the test statistic $\hat{z}$, which is asymptotically standard normal:

$$\hat{z} = \frac{\hat{\mu}_{OT}\hat{\sigma}_{IT} - \hat{\mu}_{IT}\hat{\sigma}_{OT}}{\sqrt{\hat{\theta}}} \tag{D.1}$$

where

$$\hat{\theta} = \frac{1}{N}\left(2\hat{\sigma}^2_{IT}\hat{\sigma}^2_{OT} - 2\hat{\sigma}_{IT}\hat{\sigma}_{OT}\hat{\sigma}_{IT,OT} + \frac{1}{2}\hat{\mu}^2_{OT}\hat{\sigma}^2_{IT} + \frac{1}{2}\hat{\mu}^2_{IT}\hat{\sigma}^2_{OT} - \frac{\hat{\mu}_{IT}\hat{\mu}_{OT}}{\hat{\sigma}_{IT}\hat{\sigma}_{OT}}\hat{\sigma}^2_{IT,OT}\right). \tag{D.2}$$

The test relies on the assumption that the returns are serially uncorrelated. This assumption is satisfied in our data. It also relies on the assumption that the returns are normally distributed. Whereas monthly index returns are fairly normally distributed, the returns of options are not. The problem is mitigated in a portfolio of the index, bond, and options.

For the stochastic dominance tests, the sample counterpart of conditions (7) and (8), applied to the two distributions drawn from their respective populations, is that we must have for every z in the joint support:

$$\hat{D}^2_{IT}(z) - \hat{D}^2_{OT}(z) > 0, \tag{D.3}$$

where

$$\hat{D}^2_J(z) = \frac{1}{N}\sum_{i=1}^{N}\left(z - W_{Ji}\right)_+ , \tag{D.4}$$

N is the number of paired outcomes, W_{Ji} is the i^{th} outcome of the sample J, and $(x)_+ \equiv \max(x,0)$.[28] Clearly, if (D.3) is violated at any point in the *interior* of the joint support, the null of non-dominance cannot be rejected. On the other hand, (D.3) becomes, by definition, equality at one or both endpoints of the support. The DD (2007) test deals with this problem by restricting the set of points over which (D.3)-(D.4) 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 not

[28] (D.4) is a discrete data analog for the integral formula (7) in Section II.D. See DD (2000) for further details.

significant. As opposed to DD (2007), 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:

$$\hat{T}(z) = \frac{\hat{D}_{IT}^2(z) - \hat{D}_{OT}^2(z)}{\sqrt{\hat{V}^2(z)}}, \tag{D.5}$$

where the numerator is given by (D.3)-(D.4) and

$$\hat{V}^2(z) = \hat{V}_{IT}^2(z) + \hat{V}_{OT}^2(z) - 2\hat{V}_{IT,OT}^2(z) \tag{D.6}$$

$$\hat{V}_I^2(z) = \frac{1}{N}\left[\frac{1}{N}\sum_{i=1}^{N}\left(z - W_{Ii}\right)_+^2 - \hat{D}_I^2(z)^2\right], \quad I = IT, OT \tag{D.7}$$

and

$$\hat{V}_{OT,IT}^2(z) = \frac{1}{N}\left[\frac{1}{N}\sum_{i=1}^{N}\left(z - W_{ITi}\right)_+\left(z - W_{OTi}\right)_+ - \hat{D}_{IT}^2(z)\hat{D}_{OT}^2(z)\right]. \tag{D.8}$$

The maximal and minimal values of the statistic are calculated as a maximum and minimum of (D.5) over a set of points of z, as explained below. Stoline and Ury (1979) provide tables for the non-standard 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 (2007) develop the concept of *restricted* stochastic dominance in testing the null hypothesis $H_0 : OT \not\succ_2 IT$. 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.[29] Having derived the minimal $\hat{T}(z)$-statistic in a restricted interval, the DD (2007) 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 (2007) is that condition (D.3) holds for our sample. By our trading strategies, condition (D.3) 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 (D.5)-(D.8). This statistic is computed for the values of *z* that are sample points within the restricted interval, i.e., in this interval we have coupled observations of W_{IT} and W_{OT}, transformed to

[29] 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 (D.3) 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.

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.[30] If the minimal value is negative anywhere in the full support, the null of non-dominance is accepted with p-value of 1.[31] Otherwise, we apply the bootstrap approach for the derivation of the *p*-values for the null hypothesis, as described in detail in DD (2007) 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.[32] In our tests, we use 999 bootstrap replications in order to derive the *p*-values in the tables.

[Figures 3 and 4 about here]

Figure 3 displays the $\hat{T}(z)$-statistics for call bid prices plotted against the annualized arithmetic return on the OT and IT portfolios *z*. It is clear from the figure that $\hat{T}(z)$ is initially a increasing and then decreasing function of *z* as we may expect since, by virtue of our trading strategy, OT is better off in relatively low states and worse off in relatively high states of the economy. It is also clear that the $\hat{T}(z)$-statistics are positive everywhere in the

---

[30] It may 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.

[31] In this regard, the DD (2007) 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 $H_0 : IT \succ_2 OT$ while the former test accepts the null of non-dominance.

[32] 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 that the $T(z)$ is set equal to zero at the sample point at which it attains its minimum. See Davidson (2007) for further details.

joint support for the OT and IT. Figure 4 presents several instructive cases for the DD (2007) and DD (2000) tests from our data.

There is a cost in adopting the DD (2007) null, because, as it can be analytically shown, this null cannot be rejected over the entire support of the sample distribution. DD (2007) 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 (2007) 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 this extreme belongs.

References

Baillie, R. T., and T. Bollerslev, 1992, "Prediction in Dynamic Models with Time-Dependent Conditional Variances," *Journal of Econometrics* 52, 91-113.

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

Bogle, J. C., 2005, "The Mutual Fund Industry 60 Years Later: For Better or Worse?" *Financial Analyst Journal* 61, 15-24.

Cerny, A., 2004, "Introduction to Fast Fourier Transform in Finance," *Journal of Derivatives* 12, 73-88.

Constantinides, G. M., 1986, "Capital Market Equilibrium with Transaction Costs," *Journal of Political Economy* 94, 842-862.

Constantinides, G. M., J. C. Jackwerth, and S. Perrakis, 2008, "Mispricing of S&P 500 Index Options," *Review of Financial Studies*, forthcoming.

Constantinides, G. M., and S. Perrakis, 2002, "Stochastic Dominance Bounds on Derivatives Prices in a Multiperiod Economy with Proportional Transaction Costs," *Journal of Economic Dynamics and Control* 26, 1323-1352.

Constantinides, G. M., and S. Perrakis, 2007, "Stochastic Dominance Bounds on American Option Prices in Markets with Frictions," *Review of Finance* 11, 71-115.

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

Davidson, R., 2007, "Testing for Restricted Stochastic Dominance: Some Further Results," Working Paper, McGill University.

Davidson, R., and J.-Y. Duclos, 2000, "Statistical Inference for Stochastic Dominance and for the Measurement of Poverty and Inequality," *Econometrica* 68, 1435-1464.

Davidson, R., and J.-Y. Duclos, 2007, "Testing for Restricted Stochastic Dominance," Working Paper, McGill University.

Härdle, W., 1990, *Applied Nonparametric Regression*, Cambridge Univ. Press.

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

Jackwerth, J. C., and M. Rubinstein, 1996, "Recovering Probability Distributions from Option Prices," *Journal of Finance* 51, 1611-1631.

Jobson, J. D., and B. M. Korkie, 1981, "Performance Hypothesis Testing with the Sharpe and Treynor Measures," *Journal of Finance* 4, 889-908.

McDonald, R. L., 2006, *Derivatives Markets*, Addison-Wesley.

Memmel, C., 2003, "Performance Hypothesis Testing with the Sharpe Ratio," *Finance Letters* 1, 21-23.

Nelson, M., 1991, "Conditional Heteroskedasticity in Asset Returns: A New Approach," *Econometrica* 59, 347-370.

Perrakis, S. and M. Czerwonko, 2006, "Transaction Costs and Stochastic Dominance Efficiency in the Index Futures Options Market," Working paper, Concordia University.

Rubinstein, M., 1994, "Implied Binomial Trees," *Journal of Finance* 49, 771-818.

Rubinstein, M., 1998, "Edgeworth Binomial Trees," *Journal of Derivatives* 5, 20-27.

Stoline, M. R., and H. 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 X. Zhang, 2003, “A Monte Carlo Investigation of Some Tests for Stochastic Dominance,” Working paper, Monash University.

Yu, K., and M. C. Jones, 1998, “Local Linear Quantile Regression,” *Journal of the American Statistical Association* 93, 228-237.

Table I
Prediction Error of Monthly Volatility, 1983-2006

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 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 pre-crash 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 till the option expiration date.

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

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

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

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

This 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.

Table IV
Returns of Options Trader and Index Trader

Volatility prediction mode	#months with viol. (# months)	$\frac{\hat{\mu}_{OT}}{\hat{\sigma}_{OT}} - \frac{\hat{\mu}_{IT}}{\hat{\sigma}_{IT}}$	$\hat{\mu}_{OT} - \hat{\mu}_{IT}$	DD (2000) *p*-value $H_0 : IT \succ_2 OT$	DD (2007) *p*-value $H_0 : OT \nsucc_2 IT$ without trimming in r. t.	Minimal trimming (%) in r. t. to reject $H_0 : OT \nsucc_2 IT$ at 5%	Minimal trimming (%) in r. t. to reject $H_0 : OT \nsucc_2 IT$ at 1%
			A: Call Upper Bound				
Unconditional	43 (247)	0.068***	0.0031 1)	<0.01	0.275	4.0	6.5
90-day	100 (247)	0.087***	0.0043 2)	<0.01	0.201	3.2	5.7
Adjusted IV	120 (226)	0.123***	0.0066*, 2)	<0.01	0.124	2.6	5.3
EGARCH	65 (247)	0.097***	0.0062**, 2)	<0.01	0.086	1.6	3.2
			B: Put Upper Bound				
Unconditional	23 (247)	0.026***	0.0009	>0.1	0.391	9.7	>10
90-day	16 (247)	0.004	-0.0008	>0.1	1	n/a	n/a
Adjusted IV	4 (226)	n/a	n/a	n/a	n/a	n/a	n/a
EGARCH	9 (247)	n/a	n/a	n/a	n/a	n/a	n/a

Equally weighted average of all violating options equivalent to one option per share is traded at each date. The approach of Jobson and Korkie (1981) with the Memmel (2003) correction is used to test the difference in Sharpe ratios of the *OT* and *IT* traders. The symbols *, ** and *** denote a difference in the Sharpe ratios significant at the 10%, 5% and 1% levels, respectively, in a one-sided test. The same symbols are used to denote the significance of the difference in sample means of the *OT* and *IT* traders via a one-sided bootstrap test with 9,999 trials. The symbols 1) and 2) respectively denote a positive difference in sample means of the *OT* and *IT* traders in more than 80% and exactly 100% block-bootstrap trials with paired samples of the size 120 drawn at each trial. Maximal *t*-statistics for Davidson-Duclos (DD, 2000) test are compared to critical values of Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for three 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 Stoline and Ury (1979) tables and are not reported here. P-values for the Davidson-Duclos (2007) test are based on 999 bootstrap trials. The p-values for $H_0 : IT \nsucc_2 OT$ are equal to one and are not reported here. In all cases, 10% trimming on the left tail of the joint support of the *OT* and *IT* traders' returns is applied. Successive trimming on the right tail (r. t.) of this support is applied until a given significance level reported in the last to column is reached.

Table V
Returns of Straddles/Strangles Trader and Index Trader

Volatility prediction mode	# months with viol. (# months)	$\frac{\hat{\mu}_{OT}}{\hat{\sigma}_{OT}} - \frac{\hat{\mu}_{IT}}{\hat{\sigma}_{IT}}$	$\hat{\mu}_{OT} - \hat{\mu}_{IT}$	DD (2000) p-value $H_0 : IT \succ_2 OT$	DD (2007) p-value $H_0 : OT \nsucc_2 IT$ without trimming in r. t.	Minimal trimming (%) in r. t. to reject $H_0 : OT \nsucc_2 IT$ at 5%	Minimal trimming (%) in r. t. to reject $H_0 : OT \nsucc_2 IT$ at 1%
				A: Straddles			
Unconditional	34 (247)	0.099***	0.0058	>0.1	0.300	9.7	not feasible
90-day	66 (247)	0.131***	0.0068 2)	>0.1	0.277	8.1	>10
Adjusted IV	71 (226)	0.234***	0.0165*, 2)	<0.05	0.057	0.9	1.8
EGARCH	40 (247)	0.206***	0.0158**, 2)	>0.1	0.048	0.0	2.4
				B: Strangles			
Unconditional	40 (247)	0.141***	0.0081 1)	>0.1	0.226	8.1	>10
90-day	80 (247)	0.229***	0.0143*, 2)	<0.05	0.092	4.0	5.6
Adjusted IV	94 (226)	0.335***	0.0235**, 2)	<0.05	0.020	0.0	1.8
EGARCH	54 (247)	0.233***	0.0172**, 2)	>0.1	0.020	0.0	1.6

Equally weighted average of all violating options equivalent to one call and one put per share was traded at each date. Trades were executed whenever there was a call violating the upper bound and a put traded at the same strike (for straddles) or within 0.98-1.02 moneyness bound (for strangles) for the same date. The approach of Jobson and Korkie (1981) with the Memmel (2003) correction is used to test the difference in Sharpe ratios of the *OT* and *IT* traders. The symbols *, ** and *** denote a difference in the Sharpe ratios significant at the 10%, 5% and 1% levels, respectively, in a one-sided test. The same symbols are used to denote the significance of the difference in sample means of the *OT* and *IT* traders via a one-sided bootstrap test with 9,999 trials. The symbols 1) and 2) respectively denote a positive difference in sample means of the *OT* and *IT* traders in more than 80% and exactly 100% block-bootstrap trials with paired samples of the size 120 drawn at each trial. Maximal t-statistics for Davidson-Duclos (DD, 2000) test are compared to critical values of Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for three 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 Stoline and Ury (1979) tables and are not reported here. P-values for the Davidson-Duclos (2007) test are based on 999 bootstrap trials. The p-values for $H_0 : IT \nsucc_2 OT$ are equal to one and are not reported here. In all cases 10% trimming on the left tail of the joint support of the *OT* and *IT* traders' returns is applied. Successive trimming on the right tail (r. t.) of this support is applied until a given significance level reported in the last to column is reached.

Table VI

Returns of Options Trader and Index Trader—without Futures Basis Risk

Volatility prediction mode	# months with viol. (# months)	$\frac{\hat{\mu}_{OT}}{\hat{\sigma}_{OT}} - \frac{\hat{\mu}_{IT}}{\hat{\sigma}_{IT}}$	$\hat{\mu}_{OT} - \hat{\mu}_{IT}$	DD (2000) p-value $H_0 : IT \succ_2 OT$	DD (2007) p-value $H_0 : OT \nsucc_2 IT$ without trimming in r. t.	Minimal trimming (%) in r. t. to reject $H_0 : OT \nsucc_2 IT$ at 5%	Minimal trimming (%) in r. t. to reject $H_0 : OT \nsucc_2 IT$ at 1%
			A: Call Upper Bound				
Unconditional	67 (247)	0.060***	0.0012 1)	<0.01	0.434	8.1	8.1
90-day	156 (247)	0.165***	0.0083**, 2)	<0.01	0.094	1.6	4.9
Adjusted IV	195 (226)	0.151***	0.0032	<0.01	0.348	9.7	>10
EGARCH	112 (247)	0.098***	0.0037 1)	<0.01	0.282	6.4	7.2
			B: Put Upper Bound				
Unconditional	36 (247)	0.010	-0.0015	<0.1	1	n/a	n/a
90-day	52 (247)	0.030**	-0.0003	<0.01	1	n/a	n/a
Adjusted IV	64 (226)	0.034**	-0.0012	<0.01	1	n/a	n/a
EGARCH	38 (247)	0.040***	0.0014	<0.01	0.387	8.9	9.7

The table differs from Table IV only in that the basis risk is set at zero, $\bar{\varepsilon} = 0$, instead of bounding the risk by $\bar{\varepsilon} = 0.5$. Equally weighted average of all violating options equivalent to one option per share is traded at each date. The approach of Jobson and Korkie (1981) with the Memmel (2003) correction is used to test the difference in Sharpe ratios of the *OT* and *IT* traders. The symbols *, ** and *** denote a difference in the Sharpe ratios significant at the 10%, 5% and 1% levels, respectively, in a one-sided test. The same symbols are used to denote the significance of the difference in sample means of the *OT* and *IT* traders via a one-sided bootstrap test with 9,999 trials. The symbols 1) and 2) respectively denote a positive difference in sample means of the *OT* and *IT* traders in more than 80% and exactly 100% block-bootstrap trials with paired samples of the size 120 drawn at each trial. Maximal *t*-statistics for Davidson-Duclos (DD, 2000) test are compared to critical values of Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for three 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 Stoline and Ury (1979) tables and are not reported here. P-values for the Davidson-Duclos (2007) test are based on 999 bootstrap trials. The p-values for $H_0 : IT \nsucc_2 OT$ are equal to one and are not reported here. In all cases 10% trimming on the left tail of the joint support of the *OT* and *IT* traders' returns is applied. Successive trimming on the right tail (r. t.) of this support is applied until a given significance level reported in the last to column is reached.

Table VII

Returns of Options Trader and Index Trader—with Risk Aversion Coefficient 10

Volatility prediction mode	# months with viol. (# months)	$\frac{\hat{\mu}_{OT}}{\hat{\sigma}_{OT}} - \frac{\hat{\mu}_{IT}}{\hat{\sigma}_{IT}}$	$\hat{\mu}_{OT} - \hat{\mu}_{IT}$	DD (2000) *p*-value $H_0 : IT \succ_2 OT$	DD (2007) *p*-value $H_0 : OT \nsucc_2 IT$ without trimming in r. t.	Minimal trimming (%) in r. t. to reject $H_0 : OT \nsucc_2 IT$ at 5%	Minimal trimming (%) in r. t. to reject $H_0 : OT \nsucc_2 IT$ at 1%
			A: Call Upper Bound				
Unconditional	43 (247)	0.063***	0.0006 1)	<0.01	0.275	4.0	6.5
90-day	100 (247)	0.082***	0.0008 2)	<0.01	0.201	3.2	5.7
Adjusted IV	120 (226)	0.118***	0.0013*, 2)	<0.01	0.124	2.6	5.3
EGARCH	65 (247)	0.094***	0.0012**, 2)	<0.01	0.083	1.6	3.2
			B: Put Upper Bound				
Unconditional	23 (247)	0.023**	0.0002	>0.1	0.427	9.7	not feasible
90-day	16 (247)	0.001	-0.0002	>0.1	1	n/a	n/a
Adjusted IV	4 (226)	n/a	n/a	n/a	n/a	n/a	n/a
EGARCH	9 (247)	n/a	n/a	n/a	n/a	n/a	n/a

The table differs from Table IV only in that the risk aversion coefficient is set to 10, instead of 2. Equally Equally weighted average of all violating options equivalent to one option per share is traded at each date. The approach of Jobson and Korkie (1981) with the Memmel (2003) correction is used to test the difference in Sharpe ratios of the *OT* and *IT* traders. The symbols *, ** and *** denote a difference in the Sharpe ratios significant at the 10%, 5% and 1% levels, respectively, in a one-sided test. The same symbols are used to denote the significance of the difference in sample means of the *OT* and *IT* traders via a one-sided bootstrap test with 9,999 trials. The symbols 1) and 2) respectively denote a positive difference in sample means of the *OT* and *IT* traders in more than 80% and exactly 100% block-bootstrap trials with paired samples of the size 120 drawn at each trial. Maximal *t*-statistics for Davidson-Duclos (DD, 2000) test are compared to critical values of Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for three 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 Stoline and Ury (1979) tables and are not reported here. P-values for the Davidson-Duclos (2007) test are based on 999 bootstrap trials. The p-values for $H_0 : IT \nsucc_2 OT$ are equal to one and are not reported here. In all cases 10% trimming on the left tail of the joint support of the *OT* and *IT* traders' returns is applied. Successive trimming on the right tail (r. t.) of this support is applied until a given significance level reported in the last to column is reached.

Table VIII

Returns of Options Trader and Index Trader—with Equity Risk Premium 6%

Volatility prediction mode	# months with viol. (# months)	$\frac{\hat{\mu}_{OT}}{\hat{\sigma}_{OT}} - \frac{\hat{\mu}_{IT}}{\hat{\sigma}_{IT}}$	$\hat{\mu}_{OT} - \hat{\mu}_{IT}$	DD (2000) *p*-value $H_0 : IT \succ_2 OT$	DD (2007) *p*-value $H_0 : OT \nsucc_2 IT$ without trimming in r. t.	Minimal trimming (%) in r. t. to reject $H_0 : OT \nsucc_2 IT$ at 5%	Minimal trimming (%) in r. t. to reject $H_0 : OT \nsucc_2 IT$ at 1%
			A: Call Upper Bound				
Unconditional	38 (247)	0.038***	0.0009 1)	<0.01	0.434	6.4	6.5
90-day	85 (247)	0.075***	0.0035 2)	<0.01	0.228	4.0	5.7
Adjusted IV	96 (226)	0.094***	0.0052*, 2)	<0.01	0.156	2.6	5.3
EGARCH	58 (247)	0.082***	0.0051*, 2)	<0.01	0.118	2.4	4.0
			B: Put Upper Bound				
Unconditional	23 (247)	0.024**	0.0008	>0.1	0.425	6.4	>10
90-day	11 (247)	0.009	-0.0000	>0.1	1	n/a	n/a
Adjusted IV	3 (226)	n/a	n/a	n/a	n/a	n/a	n/a
EGARCH	6 (247)	n/a	n/a	n/a	n/a	n/a	n/a

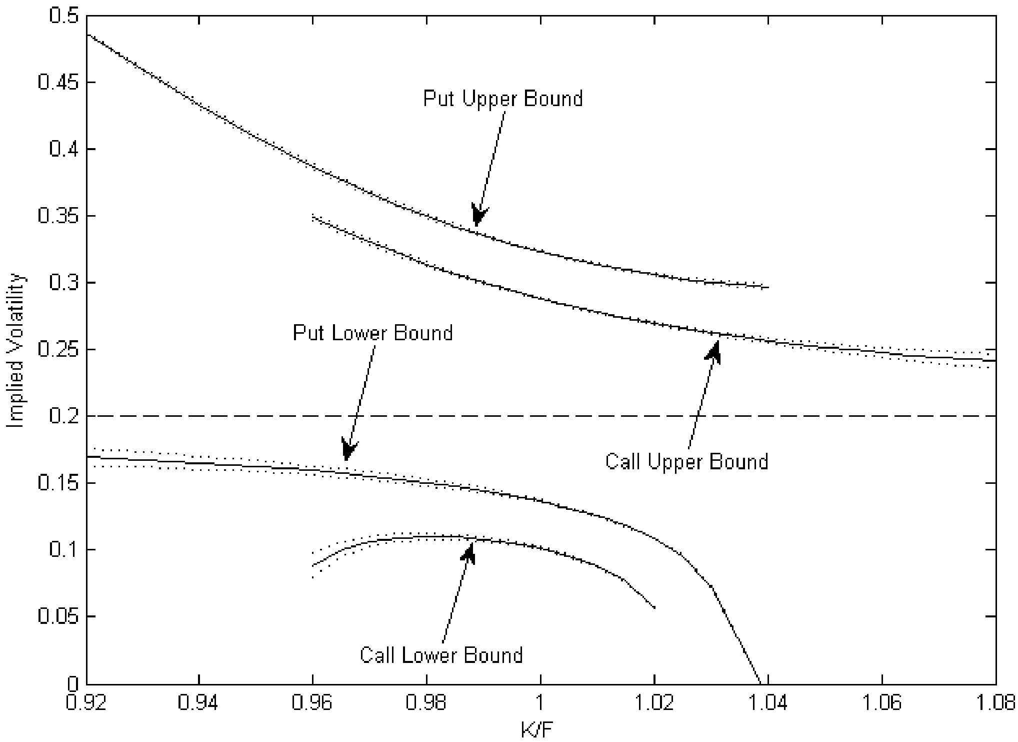
The table differs from Table IV only in that the risk premium is set to 6%, instead of 4%. Equally weighted average of all violating options equivalent to one option per share is traded at each date. The approach of Jobson and Korkie (1981) with the Memmel (2003) correction is used to test the difference in Sharpe ratios of the *OT* and *IT* traders. The symbols *, ** and *** denote a difference in the Sharpe ratios significant at the 10%, 5% and 1% levels, respectively, in a one-sided test. The same symbols are used to denote the significance of the difference in sample means of the *OT* and *IT* traders via a one-sided bootstrap test with 9,999 trials. The symbols 1) and 2) respectively denote a positive difference in sample means of the *OT* and *IT* traders in more than 80% and exactly 100% block-bootstrap trials with paired samples of the size 120 drawn at each trial. Maximal *t*-statistics for Davidson-Duclos (DD, 2000) test are compared to critical values of Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for three 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 Stoline and Ury (1979) tables and are not reported here. P-values for the Davidson-Duclos (2007) test are based on 999 bootstrap trials. The p-values for $H_0 : IT \nsucc_2 OT$ are equal to one and are not reported here. In all cases 10% trimming on the left tail of the joint support of the *OT* and *IT* traders' returns is applied. Successive trimming on the right tail (r. t.) of this support is applied until a given significance level reported in the last to column is reached.

Table IX
Returns of Naïve Trader and Index Trader

Trade type	# months with viol. (# months)	$\frac{\hat{\mu}_{OT}}{\hat{\sigma}_{OT}} - \frac{\hat{\mu}_{IT}}{\hat{\sigma}_{IT}}$	$\hat{\mu}_{OT} - \hat{\mu}_{IT}$	DD (2000) *p*-value $H_0: IT \succ_2 OT$ ($H_0: OT \succ_2 IT$)	DD (2007) *p*-value $H_0: OT \nsucc_2 IT$ without trimming in r. t.	Minimal trimming (%) in r. t. to reject $H_0: OT \nsucc_2 IT$ at 5%	Minimal trimming (%) in r. t. to reject $H_0: OT \nsucc_2 IT$ at 1%
				A: All Quantiles			
Short call	247 (247)	0.209***	0.0060 1)	<0.01	0.223	6.5	9.7
Short put	247 (247)	0.598***	0.0078	<0.01	0.266	not feasible	not feasible
Long call	247 (247)	-0.669***	-0.0403***	<0.01	1	n/a	n/a
Long put	247 (247)	-0.334***	-0.0292***	<0.01 (<0.1)	1	n/a	n/a
				B: 10th or 90th Critical Quantile			
Short call	58 (243)	0.065***	0.0041 2)	<0.01	0.194	9.1	>10
Short put	67 (243)	0.084***	0.0034 1)	<0.01	0.320	>10	>10
Long call	73 (243)	-0.158***	-0.0149***	>0.1	1	n/a	n/a
Long put	95 (243)	-0.054***	-0.0058	>0.1	1	n/a	n/a
				C: 2.5th or 97.5th Critical Quantile			
Short call	32 (243)	0.071***	0.0057 2)	<0.01	0.073	not feasible	not feasible
Short put	36 (243)	0.050***	0.0022 1)	>0.1	0.336	not feasible	not feasible
Long call	27 (243)	-0.034***	-0.0038	>0.1	1	n/a	n/a
Long put	45 (243)	0.006	+0.0000	>0.1	0.473	>10	>10

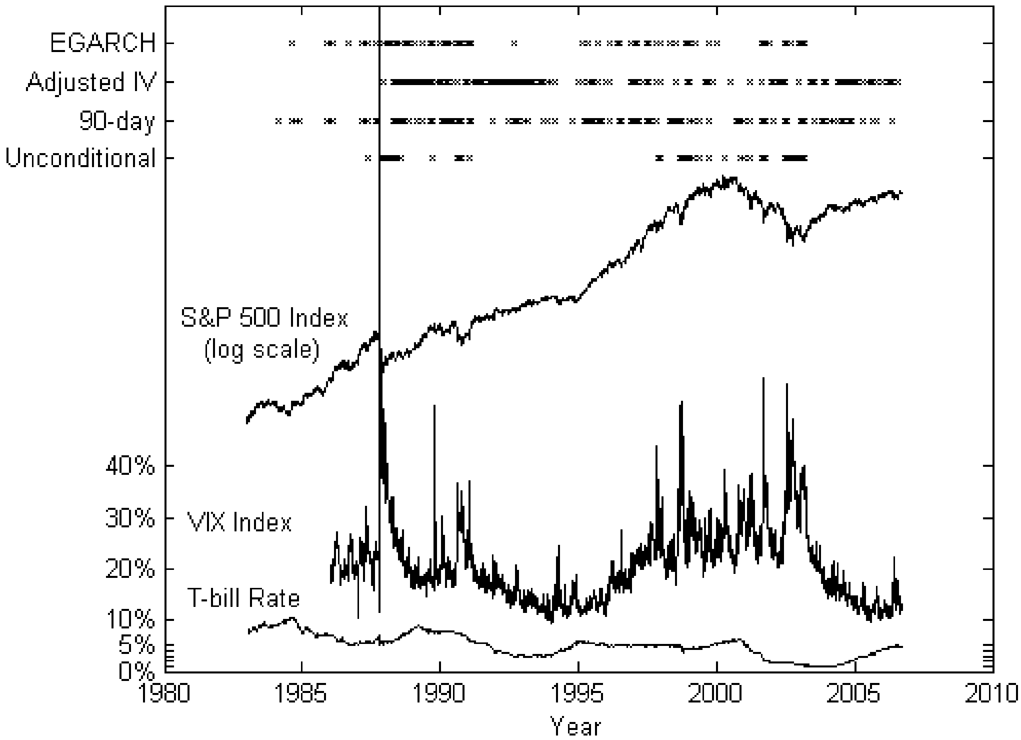
Equally weighted average of all violating options equivalent to one option per share is traded at each date. The approach of Jobson and Korkie (1981) with the Memmel (2003) correction is used to test the difference in Sharpe ratios of the *OT* and *IT* traders. The symbols *, ** and *** denote a difference in the Sharpe ratios significant at the 10%, 5% and 1% levels, respectively, in a one-sided test. The same symbols are used to denote the significance of the difference in sample means of the *OT* and *IT* traders via a one-sided bootstrap test with 9,999 trials. The symbols 1) and 2) respectively denote a positive difference in sample means of the *OT* and *IT* traders in more than 80% and exactly 100% block-bootstrap trials with paired samples of the size 120 drawn at each trial. Maximal *t*-statistics for Davidson-Duclos (DD, 2000) test are compared to critical values of Studentized Maximum Modulus Distribution tabulated in Stoline and Ury (1979) for three nominal levels of 1, 5, and 10% with $k = 20$ and $\nu = \infty$. The p-values for $H_0: OT \succ_2 IT$, which are greater than 10%, the highest nominal level available in Stoline and Ury (1979) tables are not reported here. P-values for the Davidson-Duclos (2007) test are based on 999 bootstrap trials. The p-values for $H_0: IT \nsucc_2 OT$ are equal to one and are not reported here. In all cases 10% trimming on the left tail of the joint support of the *OT* and *IT* traders' returns is applied. Successive trimming on the right tail (r. t.) of this support is applied until a given significance level reported in the last to column is reached.

Figure 1: Illustration of Upper and Lower Bounds on Call and Put Options



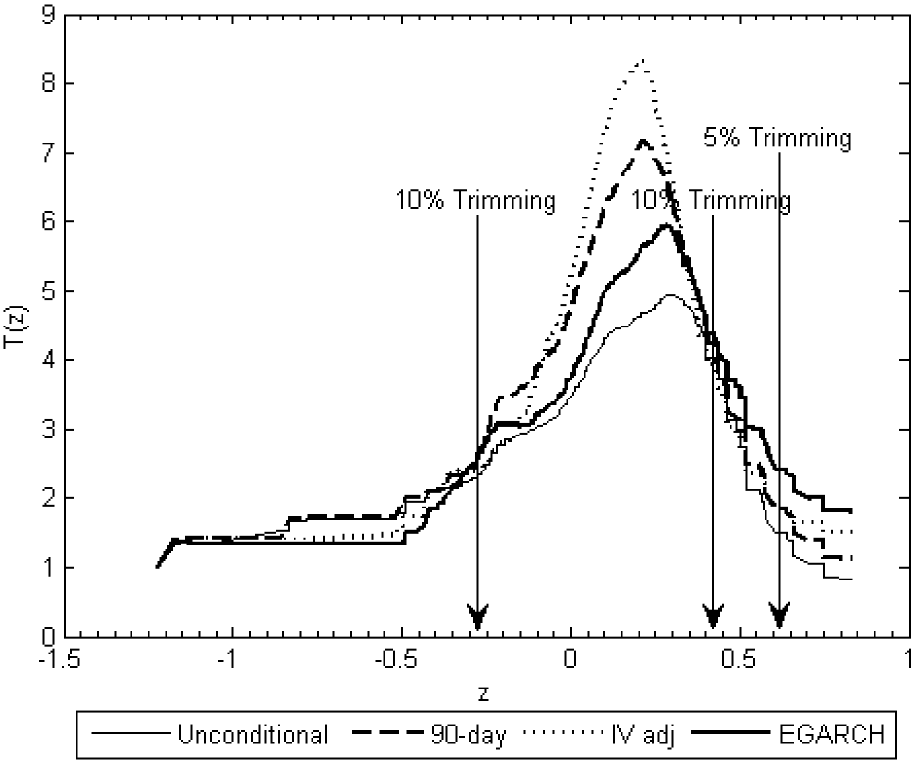
Bounds are derived for $\sigma = 0.20$ imposed on a 90-day past distribution of daily S&P 500 returns for May 22, 1996. 95% upper and lower confidence intervals represented by dotted lines are derived by bootstrapping the 90-day distribution. The results exemplify the dependence of the bounds on the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

Figure 2: Time Distribution of Observed Violations



The crosses display the violations of the call upper bound for the period February 1983-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 4th, 1986. The value for VIX just prior to the October 1987 crash was 170% and it was trimmed to facilitate presentation.

Figure 3: Sample T-statistics (Standardized Stochastic Dominance Function) for the Call Upper Bound



The figure displays the sample T-statistics (standardized second order stochastic dominance function $\hat{D}^2_{IT}(z) - \hat{D}^2_{OT}(z)$) for the call upper bound. The independent random variable z is the annualized arithmetic return on wealth liquidated at the 247 option expiration dates.

Figure 4: Examples of **s**ample T-statistics (Standardized Stochastic Dominance Function)

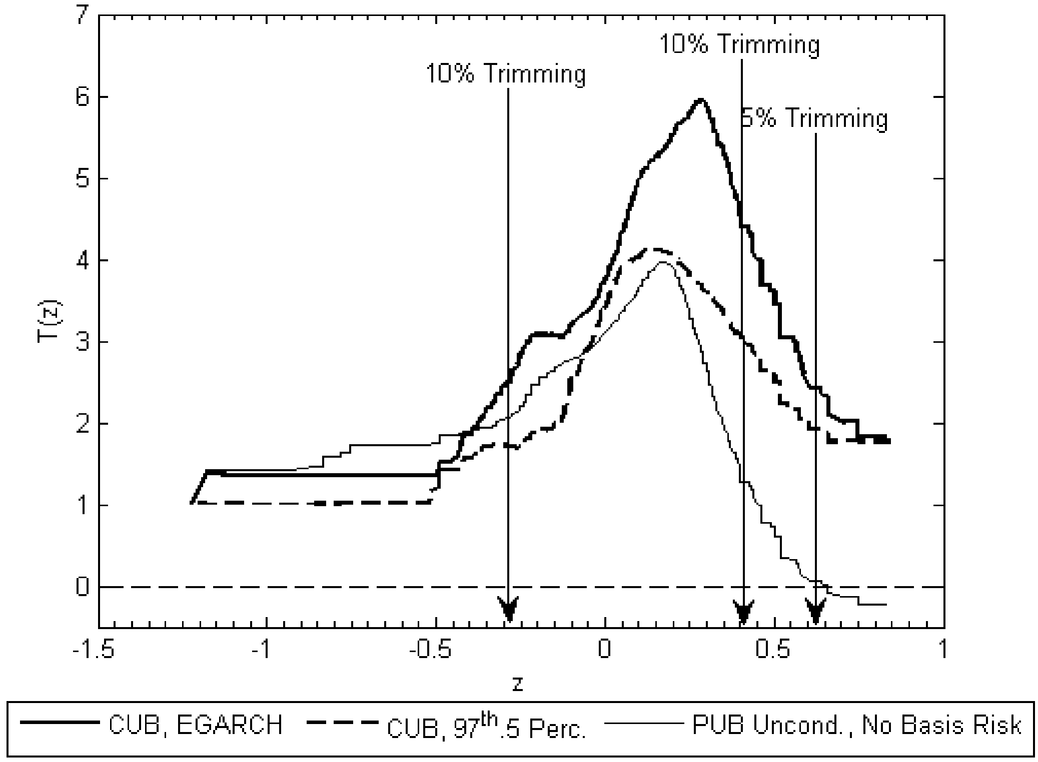


Figure 4 exemplifies various cases we subject to the DD (2007) and DD (2000) tests. It is clear in this figure that for the put upper bound without futures basis risk (PUB in the figure) we should not reject the null of nondominance since we observe negative *T*-statistics in the right tail; nevertheless the DD (2000) test provides evidence for stochastic dominance as shown in Table VI. It appears from Figure 5 that we should reject this null for the two remaining cases, the call upper bound for the EGARCH volatility estimation (CUB, EGARCH in the figure) and the call upper bound for Naïve Practitioner (CUB, 97^{th}.5 Perc. in the figure). Even without trimming in the right tail we reject in both cases with the respective *p*-values of 0.086 and 0.073 in Tables IV and IX. However, it also appears form Figure 5 that the distribution for CUB, EGARCH is clearly superior to the one for Naïve Practitioner since the former envelopes from above the latter everywhere except the limits of the joint support. To demonstrate this superiority, we need to apply trimming in the right tail, which produces rejections at 5 and 1% levels for CUB, EGARCH in Table IV with very little trimming while these significance levels are not reached in Table IV for the Naïve Practitioner. Note also that the DD (2000) fails to classify these two cases, yielding the same *p*-values for the two null hypotheses it considers in both cases. The independent random variable *z* is the annualized arithmetic return on wealth liquidated at the 247 option expiration dates.