Transaction Costs and Stochastic Dominance Efficiency in the Index Futures Options Market

George M. Constantinides[#], Stylianos Perrakis[*+] and Michal Czerwonko[*]

[#] University of Chicago and NBER

[*] Respectively, Professor and PhD student at the John Molson School of Business, Concordia University. They wish to thank Canada's Social Sciences and Humanities Research Council for financial support. Comments from Jens Jackwerth and participants at doctoral seminars in Montreal are gratefully acknowledged.

[+] Corresponding author, sperrakis@jmsb.concordia.ca, 1 (514) 848-2424 x 2963

Transaction Costs and Stochastic Dominance Efficiency in the Index Futures Options Market

Abstract

This paper examines the stochastic dominance efficiency in the presence of transaction costs for S&P 500 index futures call and put options by estimating the Constantinides-Perrakis (2006) bounds on reservation write and reservation purchase prices and then verifying whether the observed option prices satisfy them. The bounds are estimated from data on past realizations of the underlying asset and under various data-based assumptions about the investor-assumed distribution of that asset. The bounds are then compared to observed market prices and several violations are identified. The paper then derives trading strategies that exploit these violations and increase expected utility for any risk-averse investor. It develops a metric that evaluates the increase in expected utility within a certain utility class and links it with the traditional second-degree stochastic dominance criterion. Last, it demonstrates by out-of-sample tests with realized underlying asset prices that these strategies to exploit the mispricing of index futures options do indeed improve risk-adjusted returns for risk-averse investors.

Keywords: Derivatives pricing; Transaction costs; Stochastic dominance; Index options; Futures options.

I. Introduction

This paper is an examination of the opportunities to realize superior risk adjusted expected returns by trading in the S&P 500 index and in the index futures option markets in the presence of transaction costs. Such superior returns appear because of the adoption of stochastically dominating strategies that are feasible whenever observed option prices violate appropriately defined upper or lower bounds. The strategies are stochastically dominating in the conventional second-degree sense, implying that they increase the expected utility of investors possessing *any* type of increasing and concave utility function. A novel feature of our approach is the fact that the dominating strategies are dynamic and incorporate proportional transaction costs in restructuring portfolios.

In our study we identify mispriced options for given estimated probability distributions of the underlying index return. These options provide opportunities to adopt stochastically dominating trading strategies that are independent of investor utility. We then test the improvement in expected utility using specific utility functions of the constant proportional risk aversion (CPRA) type. We develop a metric to measure the utility improvement for a particular type of investors and link it to the conventional second-degree stochastic dominance criterion. The metric involves an integral condition on the two distributions of terminal portfolio return with and without the trading strategy for the mispriced option. This metric corresponds to the expected excess rate of return to option expiration attained as a result of the trading strategy. We find that the improvements in this expected return, which is utility-specific but risk-adjusted in the conventional Rothchild-Stiglitz (1970) sense, are substantial. These improvements are also verified for the mispriced options in out-of-sample tests involving the bootstrapped distributions derived from observed index returns till option expiration.

The option bounds used to identify the appropriate stochastically dominating strategies were initially derived for European options in Constantinides and Perrakis (CP, 2002). They were subsequently extended to American index options and American futures options (CP, 2006). These bounds use the entire distribution of the underlying asset as input, but are otherwise free from any assumptions about type of distribution or investor utility. The multiperiod optimization model relies on results initially derived by Constantinides (1979, 1986). To our knowledge this is the first empirical study applying option pricing models to observed option prices in the presence of transactions costs.

Stochastic dominance rules for the pairwise comparison of asset returns were initially developed by Hadar and Russell (1969) and Hanoch and Levy (1969), and subsequently extended by several other writers[1]. Although these rules present considerable theoretical interest, they have had relatively limited applications for two main reasons. First, the rules provide little guidance for portfolio selection, namely the creation of efficient portfolios from a set of assets. Second, they are rather difficult to extend to multiperiod problems.

Stochastic dominance was first applied to option pricing by Levy (1985) in a single period model that derived option bounds relying on the entire distribution of stock returns. One of these bounds had first been derived by arbitrage methods by Perrakis and Ryan (1984), while Ritchken (1985) derived essentially the same bounds as Levy by relying on capital market equilibrium arguments. Both the arbitrage and the capital equilibrium (but not the stochastic dominance)

[1] See the survey paper by Levy (1992).

methods were extended into a multiperiod context by Perrakis (1986, 1988) and Ritchken and Kuo (1988).

The CP studies were a major extension of these results, insofar as they derived results similar to those of the earlier studies by using arbitrage methods augmented by the presence of transaction costs and a recursive expected utility maximization. The derivations relied heavily on regularity conditions about multiperiod utility maximization in the presence of transaction costs originally derived by Constantinides (1979). While some of the CP (2002) results were dependent on the partition of the interval to option expiration into discrete trading subperiods, some of those results derived bounds that were *independent* of such partition. It is these latter results that were extended to American options in CP (2006). Our study focuses exclusively on two of those results[2], one of which involves upper bounds on American call futures options and the other lower bounds on American put futures options.

Most option pricing models derive their results by absence of arbitrage arguments; hence, the presence of transaction costs is a major constraint on their applicability. Indeed, it is well known since Merton (1989) that even a “small” but finite rate of proportional transaction costs on the underlying asset in any restructuring of the option replicating portfolio invalidates completely the Black-Scholes (1973) option price even when stock returns are lognormal. Similar results also hold for the discrete time binomial model, in which the derived long option price increases with the size of the binomial tree. Attempts to bypass this difficulty have not, in general been

[2] Since the options are American, the bounds clearly depend on the number of allowable exercise dates prior to option expiration; they converge to a limit even if the options can be exercised at any time.

successful[3], and most arbitrage-based models do not survive the introduction of transaction costs. Consequently, transaction costs have generally been ignored in empirical work on option prices.

Transaction costs have also been introduced into option pricing through expected utility maximization models, in the form of investor portfolios containing the underlying asset, the riskless asset, and the option. The investor is constrained into holding the option till expiration. The optimal portfolio is derived through multiperiod recursive dynamic optimization, in which there are transaction costs in restructuring the portfolio. The reservation write (purchase) price is defined as the minimum (maximum) write (purchase) price for the option that makes the investor indifferent between including and not including the option into her portfolio[4]. This approach in general produces option prices that depend on investor characteristics such as wealth and utility function parameters. For these reasons it has not been very popular.

A variant of the expected utility approach derives the reservation write and reservation purchase option prices by finding the minimum or maximum option prices that are *independent* of investor wealth and make the investor at least as well off as he would have been in the absence of the option. The optimization is taken with respect to a given class of utility functions, which are generally chosen within the CPRA class. Constantinides and Zariphopoulou (1999, 2001) have derived option bounds based on this approach. They are generally looser than the ones in CP (2002), even though they are in principle applicable to a larger set of derivatives.

The CP (2002, 2006) stochastic dominance results are methodologically somewhere in-between the arbitrage and expected utility methods. While they rely on the existence of a value

---

[3] See, for instance, Leland (1985) and Soner, Shreve and Cvitanic (1995) for the continuous time model, and Boyle and Vorst (1992), Bensaid *et al* (1992), and Perrakis and Lefoll (1997) for the binomial model.

[4] Davis *et al* (1993) were the first to attempt option pricing under the expected utility method. See also Zariphopoulou (1999) for a useful non-technical summary of this approach.

function identified with the maximized utility of an investor, they are independent of the functional form of this utility, as in arbitrage. They assume the existence of *at least one* investor holding only the underlying asset and the riskless asset and (possibly) the option. While this assumption may be restrictive for options on individual stocks, its validity in the case of index options cannot be doubted, given that fact that numerous surveys have shown that a large number of US investors follow indexing strategies in their investments.[5] The derived bounds need the *entire distribution* of underlying asset returns, which can be anything. In our empirical work we use estimated distributions from past data. Several alternative derivations of the bounds are presented, based either on historical samples of observed values, or on *forward-looking* samples; in the latter the sample on which the distribution is based is time wise coincident with the sample of observed option prices. In all cases the form of the distribution is left unspecified.

A violation of the bounds triggers a trading strategy that improves investor expected utility net of transaction costs. Although such improvements were shown theoretically to exist under all forms of utility functions and asset returns, they can only be measured for special types of utility functions and underlying asset returns. For this reason we consider investors holding a portfolio of the riskless asset and a stock index whose returns are assumed *by the investors* as being lognormally distributed with a mean and variance equal to those of the sample used in deriving the bounds. The investor utility is of the constant proportional risk aversion (CPRA) type, and the investor is assumed to maximize the utility of the discounted consumption flow over a finite or infinite horizon. To our knowledge, this is the only portfolio selection model based on expected utility and incorporating transaction costs for which closed-form solutions exist in the

[5] Bogle (2005) reports that in 2004 index funds accounted for about one third of equity fund cash inflows since 2000 and represented about one seventh of equity fund assets.

literature. The infinite horizon case has been studied by Constantinides (1986), who assumed a simple consumption policy, by Davis and Norman (1990) who relaxed that assumption but found relatively little effect on the attained levels of expected utility, and by Dumas and Luciano (1991), who also studied basically the same model, but with a different objective function, which maximized the expected utility of final wealth. This latter study assumed an infinite horizon and relied on an endogenous discount factor to obtain convergence results. The finite horizon case has also been studied by several authors, including Genotte and Jung (1994), Balduzzi and Lynch (1999) and Liu and Lowenstein (2002). These studies also used as objective functions the maximization of the expected utility of terminal wealth, of the discounted consumption flow, or of a combination of these two objectives. All these finite horizon models show that for realistically long investor horizons of ten years or more the optimal investment policy is indistinguishable from that of the corresponding infinite horizon case. For this reason we adopt in our own work the Constantinides (1986) formulation, which is both computationally simpler and imposes more stringent requirements on our utility improvements.[6]

The stochastic dominance approach to option pricing in the presence of transaction costs has also been examined empirically in Constantinides, Jackwerth and Perrakis (CJP, 2006). Although the theoretical foundations of CJP are similar to the ones underlying this paper, insofar as they both examine the stochastic dominance efficiency of the options market, both the object of study and the methodological approaches of the two papers are very different from each other. CJP examine European options on the S&P 500 index, while our study concerns American index

[6] All the portfolio selection models with transaction costs contain a no trade (NT) region, which is narrower in the models that use the utility of the flow of consumption than in those that use the utility of terminal wealth. As discussed in section V, a narrower NT region works against the utility improvements from the strategies that exploit violations of mispriced options.

futures options; it must, therefore, deal with the troublesome issues of early exercise and the cost-of-carry relationship. Further, we examine the mispricing of individual options with respect to the underlying asset in a multiperiod trading model, while CJP consider the consistency of observed option prices not only with the underlying asset but also with each other; these particular results allow only a limited number of trading periods to option expiration. Last but not least, we focus here on trading strategies designed to exploit the identified mispriced options and on the profits derived from such strategies, a topic that has not been covered in CJP.

In the next section we present the theoretical results used in our empirical work. A summary of their formulation, including some minor clarifications of the existing theory, is relegated to the appendix. Section III presents the empirical results of the estimation of the bounds under various assumptions about the underlying asset distribution and identifies the observations violating these bounds using trading data on the S&P 500 index futures options. Section IV presents a summary of the adapted Constantinides (1986) model used in the investor trading strategies to exploit mispriced options including some necessary minor extensions, and section V develops the stochastic dominance-based metric to measure the utility improvements arising from such strategies. Section VI presents the empirical results of the adoption of stochastically dominant strategies in observed violations of the option bounds in our sample. Section VII concludes.

II. Bounds on Futures Options

In this section we present the theory underlying our empirical work, by defining the option bounds as they appear in CP (2006). We consider an economy in which there is at least one investor who, before the option is introduced, holds portfolios of long or short positions in a stock with price S_t (with a natural interpretation of an index) and/or in a zero-coupon risk-free bond with return *R*, equal to one plus the riskless rate of interest per period. At *t* the investor enters with x_t dollars in the bond account and with y_t / S_t *ex dividend* shares of stock. The bond trades do not incur transaction costs, but any trades of the risky asset decrease the bond account by a proportional transaction cost $k_1(k_2)$ for buying (selling) the asset. The investor makes sequential investment decisions at discrete trading dates *t* ($t = 0,\ 1\ldots\ ,T'$), where the terminal date *T'* may be finite or infinite. At date *t*, the stock pays cash dividends $\gamma_t S_t$, where the dividend yield parameters γ_t are assumed to be deterministic and known to the investor at time zero. We assume that the support of S_t is $(0,\infty)$, that S_t follows a general distribution, and that the successive rates of return on the stock are independently distributed with conditional mean return known to the investor at time zero:

(2.1) $$R_t^S \equiv E\left[(1+\gamma_{t+1})\frac{S_{t+1}}{S_t} \mid S_t\right].$$

We define the conditional mean return with the dividend reinvested in stock, net of transaction costs:

(2.2) $$\hat{R}_t^S \equiv E\left[\left(1+\frac{\gamma_{t+1}}{1+k_1}\right)\frac{S_{t+1}}{S_t} \mid S_t\right].$$

The distinction between $\hat{R}_t^S$ and R_t^S is negligible provided the dividend yield and the transaction cost rate are of order of a few percent.

The investors' objective is to maximize the utility $E[u_{T'}(W_{T'})]$, where $W_{T'}$ denotes the investor's net worth at T',[7] an objective that realistically represents the goals of a financial institution. The utility function $u_t(.)$ is assumed to be concave and increasing and is defined for both positive and negative terminal worth, but is otherwise left unspecified.

In revising optimally her portfolio the investor increases (decreases) the stock dollar holdings from y_t to $y_t + v_t$ by decreasing (increasing) the bond account from x_t to $x_t - v_t - \max[k_1 v_t, -k_2 v_t]$. We denote by $V(x_t, y_t, t)$ the indirect (recursively maximized) utility function or value function at t of the investor who does not have an option position, given analytically in CP (2006), equations (5) and (6). The investment decision variable v_t is constrained to be measurable with respect to the information set available at t. v_t is chosen by maximizing recursively $E[V(x_{t+1}, y_{t+1}, t+1) | S_t]$, where the bond account dynamics are, in the general case:

$$x_{t+1} = \{x_t - v_t - \max[k_1 v_t, -k_2 v_t]\}R + (y_t + v_t)\frac{\gamma_{t+1} S_{t+1}}{S_t}, \quad t \le T' - 1, \tag{2.3}$$

and the stock account dynamics are:

$$y_{t+1} = (y_t + v_t) S_{t+1} / S_t . \tag{2.4}$$

If υ_t^* denotes the optimal choice of v_t then we also set

$$x_t' = x_t - \upsilon_t^* - \max\left[k_1 \upsilon_t^*, -k_2 \upsilon_t^*\right], \qquad y_t' = y_t + \upsilon_t^* . \tag{2.5}$$

At the terminal date, the stock account is liquidated. The net worth is defined as

---

[7] Alternatively, the objective may be the maximization of the discounted sum of the utilities of consumption $u_t(c_t)$ at each trading date, plus possibly the utility of terminal worth. Although the CP (2005) bounds were derived under the terminal wealth objective, they remain valid without any reformulation under the other alternative objectives.

$W_{T'} = x_{T'} + y_{T'} - \max[-k_1 y_{T'}, k_2 y_{T'}]$.

In this market we now introduce a cash-settled *futures* contract[8] with maturity $T, T \leq T'$. Since in the presence of transaction costs the cost-of-carry relation cannot be assumed to hold exactly, we assume that the futures price F_t is linked to the stock price as:

$$(2.6) \qquad F_t = \alpha_t S_t + \varepsilon_t, \quad t \leq T$$

where α_t is a time-dependent parameter and the random variables $\{\varepsilon_t\}$ are distributed independently of each other and of the stock price series $\{S_t\}$. We assume that the random variables $\{\varepsilon_t\}$ are bounded from above by some given parameter $\bar{\varepsilon}$. In a frictionless market, a standard static no-arbitrage argument yields the cost-of-carry relation

$$(2.7) \qquad \alpha_t = R^{T-t} \prod_{s=t+1}^{T} \left(1+\gamma_s\right)^{-1}, \quad \varepsilon_t \equiv 0 .$$

Next, an American, cash-settled futures call option with strike *K* and expiration date T, the same or *earlier* than the delivery date of the futures, is added to the investment opportunity set. We consider the following sequence of events. An investor enters date t with endowments x_t and y_t in the bond and *ex dividend* stock accounts, respectively, and a *short* position in a futures call option. The endowments are stated net of any cash flows that the trader has incurred at date *t* or at an earlier date in writing the call, and net of the dividend payable on the stock at time *t*. First, the investor is informed whether she has been "assigned" or not. If the investor has been "assigned", then she pays $F_t - K$ in cash and has her position in the call closed out. If the investor has been "assigned", the value of the cash account becomes $x_t - \left(F_t - K\right)$.

---

[8] Note that in the CP (2005) model the investor is not allowed to hold the futures contract, which is understood simply to be an index on which options can be issued.

In such an economy it was shown in CP (2006) that there exists an upper bound on the reservation write price of a European call futures option given by

$$(2.8) \qquad \bar{C}\left(F_t, S_t, t\right) = \frac{1+k_1}{1-k_2} \max\left[N\left(S_t, t\right),\ F_t - K\right], \quad t \le T,$$

where the function $N(S_t, t)$ is defined as follows:

$$(2.9) \qquad N(S,t) = (R_t)^{-1} E[\max\{\alpha_{t+1} S_{t+1} + \bar{\varepsilon} - K, N(S_{t+1}, t+1)\} | S_t = S]$$

for $t \le T-1$, and $N\left(S,T\right) = 0$.

Similarly, a tight lower bound on the reservation purchase price of an American futures put option was also derived in CP (2006). The cash payoff of the put exercised at time t is $K - F_t = K - (\alpha_t S_t + \varepsilon_t)$, $t \le T$, with $F_T = S_T$ whenever the futures and the option expire at the same date. Here a trader enters date t with endowments x_t and y_t in the bond and stock accounts, respectively, and a *long* position in a put futures option. The endowments x_t and y_t are net of any cash flows that the trader has incurred at date t or at an earlier date in buying the put. We stipulate that, at each date, the trader may either hold on to the put position or exercise it, but is constrained from selling it. It may well be optimal for the trader, at certain times, to sell the put rather than hold on to it or exercise it.[9] If the trader exercises the put, she receives $K - (\alpha_t S_t + \varepsilon_t)$ in cash from a trader with a short position in the put that is "assigned".

[9] The reservation purchase price of a put is derived under this constrained policy. Likewise, the reservation write price of the call option in (2.8)-(2.9) is derived under the constraint that the short position may be assigned by an option holder, but may not be closed by the short investor prior to expiration. Note, however, that if a trader increases expected utility by opening an option position at a given price and is constrained not to close out the position prior to expiration, then the trader increases expected utility even further when the constraint is removed. Thus, the relations (2.8)-(2.10) are also reservation prices for unconstrained investors.

Here the lower bound on the reservation purchase price of the American futures put option was shown to be equal to

$$(2.10)\qquad \underline{P}(F_t, S_t, t) \equiv \max\left[K - F_t,\ \frac{1-k_2}{1+k_1} M(S_t,\ t)\right],\quad t \le T,$$

where the function $M(S_t, t)$ is given by

$$(2.11)\qquad M(S,t) = (R_t)^{-1} E\left[\max\left[K - (\alpha_{t+1}S_{t+1} + \bar{\varepsilon}), M(S_{t+1}, t+1)\right] | S_t = S\right],$$

for $t \le T-1$, and $M(S,T) = 0$.

A final caveat in the derivation of the bounds concerns the satisfaction of the assumed *monotonicity condition*, which was a necessary condition for the derivation of the CP (2006) bounds. This monotonicity condition is satisfied if the following sufficient conditions (2.12a) and (2.12b) hold for (2.8)-(2.9) and (2.10)-(2.11) respectively:

$$(2.12a)\qquad [y'_{t-1} + (1+\gamma)N(S_{t-1}, t-1)/(1-k_2)]/S_{t-1} - (R/R_s)^{T-t}/(1-k_2) > 0,$$

$$(2.12b)\qquad [y'_{t-1} - M(S_{t-1}, t-1)/(1+k_1)]/S_{t-1} - (R/R_s)^{T-t}/(1+k_1) > 0.$$

From (2.5) it is clear that the satisfaction of (2.12ab) for any $t < T$ depends on the ratio $\frac{y'_{t-1}}{S_{t-1}} = N_{t-1}$, where N_t denotes the number of shares optimally held at t. While this number clearly increases as a function of the investor's initial stock position, (2.5) implies that this ratio depends also on the investor utility function via the term υ_t^*. A closed-form expression for the minimum stockholdings such that (2.12ab) are satisfied with probability 1 over the time to option expiration does not exist, but it is possible to establish such minimum stockholdings via numerical simulations for the parameter values mirrored in our data and for the value function

$V(x_t, y_t, t)$ and asset dynamics described in section IV. It can be shown[10] that a ratio of 1.5 shares per option position guarantees the satisfaction of (2.12ab) for all parameter values in our data.

In the next section we test relations (2.8)-(2.11) with data on S&P 500 index futures options.

III. Empirical Results

In this section we estimate the bounds (2.8)-(2.9) and (2.10)-(2.11) and compare them respectively to observed call futures bid and put futures ask prices. As shown in CP (2006), if the observed call bid price (put ask price) exceeds the upper bound (2.8)-(2.9) (lower bound (2.10)-(2.11)) then there exists a utility-improving strategy for *any* risk averse investor holding the S&P 500 index and the riskless bond. These strategies are explored in the next section for particular types of investors. The key issue in the empirical work is the assumed information set available to the investor in the estimation of the bounds (2.8)-(2.11).

In our empirical work we use American exercise-style S&P 500 futures call and put options maturing in 30 days, and contemporaneous intraday S&P 500 futures quotes for the underlying for the years 1990-2002 to search for violations of the bounds (2.8)-(2.9) and (2.10)-(2.11); a full description of our data base is in appendix A. In estimating the bounds we *do not* assume any particular type of distribution for the underlying asset, the S&P 500 returns. Instead, we assume that the investor's information set is drawn from empirical distributions of the observed S&P 500 daily returns. We use three alternative information sets for these distributions: (i) a set of

[10] Details of the numerical demonstration are available from the authors on request. See also similar results in CP (2006).

historical daily returns on S&P 500 since 1928, (ii) a forward-looking sample of 1990-2002 daily index returns, and (iii) a set of daily returns which occurred in the course of 30 calendar days until the maturity of each option in our sample.[11] In any one of these three cases we explicitly model the first moment of the index distribution following the arguments in Merton (1980). In particular, we use a 4% *cum dividend* risk premium[12] above the 3-month T-bill rate observed at the option sampling date, where the assumed deterministic dividend yield is derived from daily cash dividends observed until the underlying futures contract maturity[13]. Daily dividend distributions on the S&P were obtained from Standard & Poors and T-bill data was obtained from Federal Reserve Bank of St. Louis Economic Research Database (FRED®).

A key issue in the bounds estimation is the ability of the investor to detect shifts in the underlying asset distribution. There is clear evidence from Table 1 that the distribution shifted some time in the later years of our options sample, with Sample 1 covering the years 1990-1996 and Sample 2 the years 1997-2002. A similar shift is seen in the volatility *implied* by the futures options according to the Black-Scholes model. This is seen in Figure 1, which presents kernel regressions of the volatility implied[14] by the bid prices for calls and the ask prices for puts under

---

[11] The time periods of the historical and forward-looking samples are broadly similar to the ones used by CJP. Our samples, however, are used to derive the distributions of the *daily* index returns, while CJP, who study European options, need only the distribution of the entire return to option expiration.

[12] We also used an 8% risk premium.

[13] Considering a dividend yield on the index ignores the lumpiness of the dividend process. The relaxation of this assumption, however, did not bring significant changes to our results.

[14] To derive the implied volatility throughout this paper we use Kamrad and Ritchken (1991) trinomial model under the perfect cost-of-carry relationship. A trinomial lattice clearly outperforms the binomial with respect to the

a Black-Scholes set of assumptions as a function of the degree of moneyness of the options for Sample 1 and Sample 2. Although the implied volatility is a meaningless metric outside the Black-Scholes assumptions, it is also a widely used statistic that investors may use to detect shifts in the distribution of the underlying asset returns. For this reason we maintained this division of the entire sample of options into two in all the empirical tests that were done.[15]

(Table 1 and Figure 1 about here)

For each one of the two option samples we estimate the bounds (2.8)-(2.11) under three alternative estimates of the underlying return distribution constructed from histograms of the following sets of observed S&P 500 daily returns. The *historical* set included all daily returns since 1928 until the beginning date of each sample. The *forward-looking* set contained all daily returns on the index for the respective subperiod. Last, the *realized* set was composed of the daily returns that actually occurred in the course of the 30 calendar days until the maturity of each option in our sample. Option bounds constructed from distributions drawn from this last set clearly correspond to the "best possible information" about the underlying return distribution. The characteristics of the three data sets are shown in Tables 1 and 2.

(Table 2 about here)

---

oscillating behavior around the convergence price, which results in far better implied volatility estimates for the former model.

[15] Note that the sample split is rather rough; for instance, for 1990 and 1991, we observed distinctively higher implied volatility than in the remaining years for Sample 1. However, since by the forward-looking distributions we proxy for an information set available to the investor, we avoid a finer sample split to keep the information set noisy in the context of a 30-day trading horizon.

Appendix A describes in detail the estimation of the various variables that enter into the bounds relations (2.8)-(2.11). We assume one-way transaction cost rates of trading in the index $k_1 = k_2 = 0.5\%$, rates that are considered representative of the costs faced by individual traders.[16] The derivation of the bounds requires recursive estimation of conditional expectations in (2.9) and (2.11), which is straightforward under the assumption of lognormality of returns, since these expressions can be derived through an adapted binomial or trinomial model. These methods are, however, clearly unsuitable here since we use (discrete) empirical distributions based on the observed daily returns on the S&P 500 index in order to estimate the bounds. Such empirical distributions must contain a large number of states in order to be realistic. To deal with this problem we develop a multinomial lattice model that aggregates similar states at each time step, which represents one trading day; a summary of this lattice model is presented in Appendix B. While this lattice model is sufficient for the quarterly options in our sample, for which exercise takes place only at the end of the day, it does not account for the higher than daily exercise frequency of serial options. To account for possibly higher bounds because of more frequent exercise opportunities we use a trinomial model for intraday trading with Black-Scholes assumptions regarding the return process using the volatility of a given set of daily index returns while holding the remaining assumptions unchanged. We compute the bound value for an option exercisable (i) once a trading day; (ii) several (eight) times a day. The difference between (ii) and (i) will serve as a proxy for the premium for continuous exercise opportunities, and it will augment the bound value for serial options. Our results indicate, however, that this premium is negligible for calls and quite low for puts, i.e. it remains within 2% of the total value of the put lower bound.

[16] See the comments in Balduzzi and Lynch (1999, p. 63).

Table 3 presents the summary results for the bounds' violations according to the three types of empirical distribution of the S&P 500 index described in the previous section. In addition, we include similar results for bounds estimated under the assumption of lognormal distributions for the index, with volatility equal to that of the corresponding empirical distribution. Figure 2 similarly presents the results for the bounds estimations and the observed option prices in terms of their implied volatilities, the latter viewed simply as a translation device. To aggregate the data, we apply kernel regression to the bounds' implied volatilities[17] as functions of the option moneyness. To derive the kernel bandwidth, we experimented around the Silverman (1986) rule of thumb, which is $0.79QN^{-1/5}$, where Q is the inter-quartile range and used this quantity increased by a factor of two.

(Table 3 and Figure 2 about here)

The results for Table 3 and Figure 2 show clearly that the bounds derived from the historical distribution cannot identify trades leading to improvements in expected utility. Both the upper bound for the call bid price and the lower bound for the put ask price are too high with respect to the observed market prices in Sample 1, leading to very few call bound violations and a very large number of put bound violations. In Sample 2 we have the reverse though less apparent a situation: there are many call upper bound violations while there are virtually no put lower bound violations. These findings lead to a conclusion that in our sample the historical distribution is not the appropriate information set to search for market option prices implying stochastic dominance violations. Therefore, in what follows, we mostly focus on the forward-looking and realized distributions.

[17] As before, we use a trinomial model with Black-Scholes type of assumptions to derive implied volatility.

The estimation of the bounds with the use of the forward-looking distribution identified approximately each fifth and tenth call bid price violating the upper bound in, respectively, Sample 1 and 2. The pattern for the put ask price is reversed, with few violations in Sample 1 and 7.4% of violations in Sample 2. These proportions of violations for calls increase significantly in both samples when the bounds are estimated with the strongest information set, which is the distribution of returns realized till the maturity of an option; this increase is less apparent for puts. The large number of violations of these bounds, which are estimated with the "best possible information" about the underlying return distribution, raises questions about the ability of the market to forecast accurately market returns, an ability that is taken for granted in many empirical studies of option pricing.[18]

A more detailed analysis of the patterns of the violations shows clearly that most options violating the bounds were out-of-the money (OTM), for both calls and puts. This is shown very clearly in Figure 3, which plots the size of the violations relative to the corresponding bound as a function of the degree of moneyness of the option.[19] The size of the violations was very large for call options, with the median violating quote equal to 1.245 of the upper bound, while for puts the same median quote was only 0.948 of the lower bound. Note also that the large proportion of dates in which violations were observed masks significant clustering that was observed in the data, especially for puts. Of the 1133 violations for call options observed in 39 days out of a total of 141 there were three dates accounting for 371 violations, 32% of the total. For put options the clustering is much stronger, with three days accounted for 131 violations, or 62% of the 211

[18] Thus, in the large literature on the volatility smile reviewed in Jackwerth (2004) it is frequently assumed that the implied volatility extracted from observed option prices is a good estimate of the "true" volatility of the underlying asset.

[19] This reflects to some extent the fact that most option quotes in our samples were OTM.

violations observed in 16 days out of a total of 140. By contrast, there were no significant differences in the violations between quarterly and serial options in both Sample 1 and Sample 2.

(Figure 3 about here)

The use of an 8% risk premium leads, as expected, to a decrease in the proportions of violating option prices, mostly within a range of 25 to 50%. This decrease is quite large, which is somewhat surprising in spite of the large increase in the parameter: the expected return in (2.8)-(2.11) plays the role of the discount rate in risk-neutral option prices, and such prices are relatively insensitive to it. The impact of the size of the risk premium on the size of the bounds is distribution-dependent and is much higher for the less volatile forward Sample 1 distribution than for Sample 2; the impact is also stronger for OTM than for at-the-money (ATM) options. Thus, raising the premium from 4% to 8% increases (decreases) the call (put) bounds by a factor of 1.09 (0.9) for a 0.95 moneyness in the forward Sample 2, and by a factor of 1.06 (0.93) for ATM options; the corresponding figures for Sample 1 are 1.23 (0.8) for OTM and 1.11 (0.89) for ATM.

More surprisingly, the results of Table 3 show that assuming a lognormal underlying return distribution has a comparatively small effect in most cases: the proportion of violations changes very little in all cases except for the forward-looking distribution in Sample 2. This is surprising because our underlying returns data presents many violations of lognormality, verified by the Kolmogorov D-test. This robustness of the violations when lognormal returns are assumed lends further strength to our utility-improving strategies described in the following section. These strategies are applied to the bounds violations identified with the forward-looking distributions and the utility improvements are verified with both in- and out-of-sample tests in section VI.

IV. Utility-Improving Strategies by Trading on Mispriced Options

In this section we develop a model of a particular investor with CPRA preferences in order to measure utility improvements arising out of trading strategies to exploit the mispriced options identified in the previous section. The model is an adaptation of the one developed by Constantinides (1986), shown in Appendix C, in which the investor maximizes the expectation of the discounted sum of the utilities of consumption $u(c_t) = \frac{c_t^\delta}{\delta}$, $\delta < 1$ over an infinite horizon.[20] Consumption c_t comes out of the riskless asset and is of the *simple* type, implying that it is a constant proportion β of the riskless asset x_t. The distribution of the risky asset $\frac{S_{t+1}}{S_t}$, which was left unspecified in the estimation of the bounds, is assumed by the investor to be lognormal with mean μ and volatility σ; in our numerical work these parameters are set equal to those of the assumed "true" data generating process. Further, as argued in Appendix C we set the stream of dividends $y_t \gamma_t \frac{S_{t+1}}{S_t}$ accruing to the bond account approximately equal to $x_t \gamma_t \lambda^* \equiv x_t h$, where $\lambda *$ is equal to $\lambda^* = \left[\frac{\mu + \gamma - r}{(1-\delta)\sigma^2}\right]\left[1 - \frac{\mu + \gamma - r}{(1-\delta)\sigma^2}\right]^{-1}$. In such a case (2.3) becomes

$$\text{(2.3')} \qquad x_{t+1} = \{x_t - v_t - \max[k_1 v_t, -k_2 v_t]\} R(1-\beta)(1+h),$$

and the asset dynamics (2.3) and (2.4) become, for v_t=0 and for Δt denoting the length of the time partition:

$$\text{(4.1)} \qquad x_{t+1} = x_t[1 + (r - \beta + h)\Delta t + o(\Delta t)], \quad y_{t+1} = y_t[1 + \mu\Delta t + \sigma\varepsilon\sqrt{\Delta t}].$$

[20] As noted in section I, the model is unchanged if a finite but realistically large horizon is assumed.

Here r is the riskless rate, μ is the ex-dividend expected return to the risky asset and σ is a volatility parameter, with the random variable $\varepsilon \sim N(0,1)$. As $\Delta t \to 0$ (4.1) tends to the asset dynamics of the Constantinides (1986) model adapted for dividends

$$(4.2) \qquad \begin{aligned} dx &= (r - \beta + h)x dt \\ dy &= \mu y dt + \sigma y d\omega \ . \end{aligned},$$

where ω is a Wiener process.

As shown in Constantinides (1986) the optimal portfolio revision policy in his model contains a *no trading* (NT) region, expressed by two limits $\underline{\lambda}$ and $\overline{\lambda}$ on the proportion y_t / x_t. The value function of the investor without a position in the derivative (the *V-investor*) then becomes

$$(4.3) \qquad V[x, y; \beta, \underline{\lambda}, \overline{\lambda}, t] \equiv Max\{E_t \int_t^\infty e^{-\rho t} u(c(s))ds\} = Max\{E_t \int_t^\infty e^{-\rho s} \delta^{-1} c^\delta (s) ds\},$$

where the maximization is over $c(s)$, the consumption at s, $\beta \equiv c / x$ is the parameter of the simple consumption policy, ρ is the time discount factor which is assumed equal to the continuous time riskless rate r, and E_t is the current time expectation over the Wiener process ω in the dynamic equations (4.2). With the addition of dividends it is also shown in the appendix that the value function V defined in (4.3) is equal to the function V^γ given by the following expression, which is an adaptation of the equivalent expression in Constantinides (1986) in the absence of dividends

$$(4.4) \qquad V^\gamma (x, y; \beta, \underline{\lambda}, \overline{\lambda}) = \frac{\beta^\delta}{\rho - \delta(r - \beta + h)} (x^\delta / \delta + A_1 x^{\delta - s_1} y^{s_1} + A_2 x^{\delta - s_2} y^{s_2}),$$

where (s_1, s_2) are the roots of:

$$(4.5) \qquad f(s) \equiv (\sigma^2 / 2)s^2 + (\mu - \sigma^2 / 2 - r + \beta - h)s - [\rho - \delta(r - \beta + h)] = 0,$$

and (A_1, A_2) are free parameters derived from substituting the value function (4.4) into the following boundary conditions (4.6ab):

(4.6a) $$(1+k)V_x = V_y\,, \quad y/x \le \underline{\lambda}\,,$$

and

(4.6b) $$(1-k)V_x = V_y,\ \ y/x \ge \overline{\lambda}\,.$$

These yield the following equations for (A_1, A_2)

(4.7a) $$(1+k)[1 + A_1(\delta - s_1)\underline{\lambda}^{s_1} + A_2(\delta - s_2)\underline{\lambda}^{s_2}] = A_1 s_1 \underline{\lambda}^{s_1 - 1} + A_2 s_2 \underline{\lambda}^{s_2 - 1}$$

and

(4.7b) $$(1-k)[1 + A_1(\delta - s_1)\overline{\lambda}^{s_1} + A_2(\delta - s_2)\overline{\lambda}^{s_2}] = A_1 s_1 \overline{\lambda}^{s_1 - 1} + A_2 s_2 \overline{\lambda}^{s_2 - 1}.$$

Next we maximize the value function (4.4) with respect to the three unknown parameters $(\beta, \underline{\lambda}, \overline{\lambda})$ and find the values of these parameters by solving the resulting first order conditions, taking also into account the boundary conditions (4.6ab). The equations are solved numerically in the empirical work. The maximized value function (4.4) is then used as the value function $V(x_t,\, y_t,\, t)$ that we use in evaluating the improvement in expected utility as a result of adopting the arbitrage strategy.

The assets in the investor portfolio follow the discrete time dynamic equations (2.3') and (2.4) without adjustment *whenever the ratio* y_t / x_t *lies inside the NT region* $(\underline{\lambda},\ \overline{\lambda})$. Outside the NT region the optimal policy was shown by Constantinides (1986) to be of the simple type, corresponding to investing to the nearest border of the NT region. Analytically, by combining (2.3') and (2.4) with the simple investment policy we have:

$$
(4.8)\qquad
\begin{array}{ll}
(y_t + v_t)/[x_t(1+h)(1-\beta) - v_t(1+k_1)] = \underline{\lambda}\ , & y_t / x_t < \underline{\lambda}\ , \\
v_t = 0\ , & y_t / x_t \in [\underline{\lambda}\ , \overline{\lambda}], \\
(y_t + v_t)/[x_t(1+h)(1-\beta) + v_t(1-k_2)] = \overline{\lambda}\ , & y_t / x_t > \overline{\lambda}\ .
\end{array}
$$

Once the parameters $(\beta, \underline{\lambda}, \overline{\lambda})$ have been estimated we can simulate a path of optimal portfolio revisions for a *V*-investor along a path of randomly generated values of S_t, $t =$ 1,... ,T. At time t = 0, we form a portfolio containing N_0 shares of the index, with N_0 set at 1.5 in our numerical work in order to satisfy the relations (2.12ab). The dollar value of the riskless asset at t = 0 is determined by fixing the initial position of the investor in the NT region; for instance, we set $y_0 / x_0 = \overline{\lambda}$ ($y_0 / x_0 = \underline{\lambda}$) for the portfolio formed at the upper (lower) boundary of the NT region. The boundaries of the NT region remain unchanged throughout the life of an option. Dividends accrue to the bond account at the rate γ per period and the investor consumes at the rate β per period from the bond account. Immediately before a revision we have[21]: $y_t = N_{t-1} S_t$ and $x_t = (1-\beta)(R x_{t-1} + N_{t\text{-}1} \gamma S_t / S_{t-1})$, $t =$ 1,... ,T. Let x'_t and y'_t denote the dollar value of, respectively, the bond and stock accounts immediately after the adjustments due to the simple investment policy have taken place. For each revision opportunity, at the prevailing index price, the *V*-investor makes the following adjustments to her stockholdings from (4.8):

i. If $y_t / x_t > \overline{\lambda}$:

$$
(4.9a)\qquad
\begin{array}{l}
\Delta N_t = v_t / S_t = (y_t - \overline{\lambda} x_t)/(S_t[1 + \overline{\lambda}(1-k_2)])\ , \\
N_t = N_{t-1} + \Delta N_t, \\
y'_t = N_t S_t\ , \\
x'_t = x_t + \Delta N_t S_t (1-k_2)\ .
\end{array}
$$

ii. If $y_t / x_t < \underline{\lambda}$:

[21] This formulation implies that consumption takes place right after receiving the interest and cash dividends.

$$
\begin{aligned}
&\Delta N_t = v_t / S_t = (\underline{\lambda} x_t - y_t)/(S_t\,[1+\underline{\lambda}(1+k_1)])\ , \\
&N_t = N_{t-1} + \Delta N_t\ , \\
&y'_t = N_t S_t\ , \\
&x'_t = x_t + \Delta N_t S_t (1+k_1)\ .
\end{aligned} \tag{4.9b}
$$

iii. If $\overline{\lambda} \geq y_t / x_t \geq \underline{\lambda}$ the investor does not trade.

Suppose now that at some time $t<T$ an observed call option bid price C exceeds the value given in (2.8)-(2.9). Then CP (2006) show that a utility-improving strategy for the *V*-investor is to write a call option and add the proceeds to the cash account. The investor becomes thus an investor with an open position in a derivative, a *J-investor*. If the immediate exercise value F_t - K exceeds the value $N(S_t, t)$ given in (2.9) the investor is assigned;[22] otherwise, the proceeds are transferred to the stock account and, as described in CP (2006), her utility would improve, since

$$J(x,\ y\ +\ C/(1+k_1),\ t)\ \geq\ V(x,\ y,\ t). \tag{4.10}$$

The *J*-investor will be assumed to follow *the same* simple portfolio revision policy as the *V*-investor, with the difference that the *J*-investor would be assigned if at any time $t+i \in (t,T)$ the immediate exercise value F_{t+i} - K were to exceed the value $N(S_{t+i}, t+i)$. Once the *J*-investor is assigned she becomes automatically a *V*-investor, with the exercise proceeds subtracted from the cash account.

Similarly, if a put option ask price P lies below the value given in (2.10)-(2.11) the investor purchases the option from the cash account, thus becoming a *J*-investor. If the option price is less

[22] Note that the assignment policy built into (2.8)-(2.9) is the *least beneficial* for the short option holder; see the definition of the *J*-function in equation (25) of CP (2006). Any other assignment would increase even more the utility of the short position.

than the immediate exercise value $K - F_t$ the option is exercised; otherwise the cash account is restored to its original value by selling stock from the stock account, in which case her utility would improve, since

(4.11) $$J(x,\ y - P/(1-k_2),\ t) \ \geq\ V(x,\ y,\ t).$$

Here again the *J*-investor will be assumed to follow *the same* simple portfolio revision policy as the *V*-investor, with the difference that the *J*-investor closes her position by exercising the option if at any time $t+i \in (t,T)$ the immediate exercise value $K - F_{t+i}$ exceeds the value $\frac{1-k_2}{1+k_1} M(S_{t+i}, t+i)$. The *J*-investor becomes thus a *V*-investor by adding the immediate exercise value to the cash account. In other words, the *J*-investor with the open position in the derivative revises her portfolio according to (4.9ab) for the same $(\beta, \underline{\lambda}, \overline{\lambda})$ parameters as the *V*-investor. Clearly, if there is a utility improvement with this constrained policy there will also be an improvement if the *J*-investor adopts an unconstrained utility maximizing portfolio revision policy.

In the next section we demonstrate that the utility improvements of the *J*- over the *V*-investor also correspond to conventional second order stochastic dominance (SSD) of the liquidating portfolio wealth at option expiration *T* of the *J*-investor over the *V*-investor. This SSD criterion is then applied to assess the utility improvements for the bounds violations identified in section III for both in-sample and out-of-sample tests.

V. Bounds Violations and Stochastic Dominance

The traditional stochastic dominance criterion refers to a pairwise comparison of two distributions, loosely identified with the returns of alternative investment strategies. Let X and Y be two random variables whose cumulative distributions are respectively F and G. Then X exhibits SSD over Y if and only if the following relation holds[23]

(5.1) $$H(z) \equiv \int_{-\infty}^{z} (G(t) - F(t))dt \geq 0$$

for all values z within the domain of the two distributions. If (5.1) holds then the expected utility of any risk averse investor is higher if she chooses X rather than Y. In other words, $E[u(X)] \geq E[u(Y)]$ for all increasing and concave functions $u(.)$. Further, $H(+\infty)$ can be easily shown to be equal to the difference in expectations $E[X] - E[Y]$.[24] This SSD criterion has had only limited applications in portfolio selection problems, since it is a single period comparison, which makes it rather unsuitable for dynamic portfolio selection problems.

It is, nonetheless, possible to use this criterion in our case in order to assess the utility improvements in adopting the appropriate trading strategy whenever the option bounds (2.8)-(2.9) or (2.10)-(2.11) are violated. We demonstrate in this section a transformation of the relations (4.10)- (4.11), the larger expected utility of the *J*- over the *V*-investor when the derivative price violates the respective bounds (2.8)-(2.9) and (2.10)-(2.11). This transformation

---

[23] See, for instance, Hadar and Russell (1969) and Hanoch and Levy (1969).

[24] Note that by definition the random variable X is *less risky* than Y in the Rothschild-Stiglitz (1970) sense, since its distribution exhibits SSD over that of Y. Hence, if we were to adjust the difference in expectation by a risk factor (for instance by computing the Sharpe ratio) the adjustment would, if anything, magnify the difference in unadjusted expectations.

compares the cumulative portfolio returns to option expiration of the *J*- and the *V*-investors in a relation similar to (5.1). Given now that (5.1) is satisfied for all z, we use (5.1) for $z = +\infty$ in order to assess the improvements in expected returns for particular trading strategies. The procedure will be illustrated in detail for trading strategies adopted when there is a violation of the call upper bound (2.8)-(2.9), with the procedure for exploiting violations of the put lower bound (2.10)-(2.11) treated as an extension.

We consider two investors of the *V*- and *J*-type who start with the same initial portfolio (x_0, y_0) containing N_0 shares of stock and liquidate their portfolios at option expiration time *T*. At some time $\tau \in [1, T]$ the *J*-investor is assigned and her cash account is reduced by the amount F_τ - K . Let R_{kT} and N_{kT}, $k = v, j$, denote, respectively, the holding period returns and the number of shares held at the call expiration *T*, with the subscripts *v* and *j* indicating the *V*- and *J*-investors. In order to derive a single measure of portfolio return that includes the transaction costs and makes R_{vT} comparable to R_{jT}, we compute the returns by applying the liquidating values of the *V*- and *J*-portfolios at time zero[25] and at the option expiration *T*, inclusive of dividends and of any intermediate assignment of the *J*-investor. Let λ denote the y/x, the stock to bond proportion right before the zero-net-cost position has been adopted by the *J*-investor at time zero, and let

(5.1) $$n_0 \equiv C_0 / S_0(1+k_1)$$

denote the number of extra shares acquired by the *J*-investor with the proceeds from writing one call option at a price of C_0. Then we have:

[25] Alternatively, we can use the acquisition value at time zero with identical results.

$$R_{vT}=\frac{x_0R-\sum_t\{v_t-\max[k_1v_t,-k_2v_t]\}R^{T-t}+D_v+N_{vT}S_T(1-k_2)}{N_0S_0(1-k_2+1/\lambda)}\ ,$$

$$R_{jT}=\frac{x_0R-\sum_t\{j_t-\max[k_1j_t,-k_2j_t]\}R^{T-t}+D_j+N_{jT}S_T(1-k_2)-\{[F_\tau-K]^+\}R^{T-\tau}}{N_0S_0(1-k_2+1/\lambda)}\ ,$$

(5.3)

$$x_0=N_0S_0/\lambda\ ,$$
$$D_k=\sum_t N_{kt}d_TR^{T-t}\ ,$$
$$N_{vT}=N_0+\sum_t v_t/S_t\ ,$$
$$N_{jT}=N_0+n_0\ +\sum_t j_t/S_t,\ \ k=v,j,$$

where the decision variables v_t and j_t are determined in (4.8) and (4.9ab), R is the riskless return observed at $t = 0$ for the expiration time T, d_t denotes an equal in each trading period dividend payment per one share, and D_k is the dollar value at T of dividends arriving in a portfolio throughout the life of an option. The first term in the numerator of the expressions for R_{kT} in (5.3) above implies that we measure the value at T of the entire portfolio transformed into bond holdings at time zero. The next two terms, however, subtract (add) the values *at T* of any changes due to the bond account dynamics incurred in the course of the life of an option.

Similar relations also hold when a trading strategy is adopted in the case of a violation of the put option lower bound (2.10)-(2.11), which are similar to (5.3) and are not repeated here for the sake of brevity. In such a case the amount of extra shares n_0 in the trading strategy of purchasing a put option is *subtracted* from N_0 and instead of (5.2) we have $n_0=P_0/S_0(1-k_2)$, where P_0 is the put option ask price. Similarly, the last term in the numerator of R_{jT} in (5.3) is replaced by the payoff of the put option, $[K-F_\tau]^+$ which is now added rather than subtracted.

The random variables R_{vT} and R_{jT} as given in (5.3) are the counterparts of the variables *Y* and *X*, whose distributions *G* and *F* enter into the SSD relation (5.1). They embody three different random effects, the assignment time τ and the associated stock price S_τ, as well as the error term ε_τ entering into the cost of carry relation (2.5) through the futures price F_τ. While no closed-form expressions can be derived for the distributions of R_{vT} and R_{jT}, it is possible to estimate them empirically by Monte Carlo simulations. In these simulations the *V*- and *J*-investors *assume* lognormal index returns as in section IV and revise their portfolios accordingly, but the index evolves according to the empirically derived distributions used to evaluate the bounds in section III, as does the error term in (2.8). The estimated distributions of R_{vT} and R_{jT} are then substituted into (5.1) in order to verify the SSD relation. This is done in the next section.

VI. Bounds Violations and Utility Improvements: Empirical Results

The first step in measuring the utility improvements of trading in mispriced options is the estimation of the NT region according to the Constantinides (1986) model, with the characteristics of the corresponding distribution used in estimating the bounds whose violations we have observed. In addition to the two first moments of the distribution, the NT region also depends on the risk aversion parameter δ. We use three values of δ, -3, -5 and -10. Table 4 shows the NT region for all the relevant cases.

(Table 4 about here)

We consider a starting position of the investor at some point within the NT region. Given an observed bound violation, the utility improvement of the *J*- over the *V*-investor will depend on the number of shares N_0 at time zero *per one zero-net-cost position* strategy represented by the writing of one call option (purchase of one put option) and increasing (decreasing) the bond account. In our simulations and tests we keep this number equal to 1.5 in order to satisfy the monotonicity sufficient conditions (2.12ab). Given this constraint, the size of the initial portfolio (x_0, y_0) is uniquely determined by the position within the NT region. In our tests we vary this position within the NT region, from $y_0 / x_0 = \underline{\lambda}$ to $y_0 / x_0 = \overline{\lambda}$ for all risk aversion parameters. Clearly, the highest utility improvements will be observed for portfolios with the highest proportion in the risky asset; these will be the portfolios for δ = -3 and for $y_0 / x_0 = \overline{\lambda}$.[26]

For each starting portfolio within the NT region we measure the utility improvement of the *J*-over the *V*-investor by simulating paths of the assets in the investor portfolio till the option expiration. The asset dynamics follow the relations (2.3') and (2.4), with the ex-dividend return $\frac{S_{t+1}}{S_t}$ sampled from the distribution used in the estimation of the bound that was violated, and the portfolio revision parameter v_t calculated from (4.8)-(4.9ab) given the NT region $(\underline{\lambda}, \overline{\lambda})$. For each such simulated path we note the corresponding returns R_{vT} and R_{jT} of the *V*- and *J*-investors from (5.3), and for a sufficiently large number of such paths we construct the distribution functions $G_k, k = v, j$ of R_{vT} and R_{jT}. We then apply the SSD criterion (5.1) to the two distributions $G_k, k = v, j$, by estimating the following relation

(6.1) $$H(z_0) \equiv \int_{-\infty}^{z_0} (G_v - G_j) dz$$

for all z_0 within the domain of the distributions. As discussed in the previous section, $H(\infty)$ (alternatively, the value of H at which $G_v = G_j = 1$) is the expected excess return across all the

[26] Note that the initial portfolios will not be of the same size.

return states. For the *J*-investor portfolio to show SSD over the *V*-investor portfolio $H(z_0)$ must be nonnegative for all z_0. A sufficient condition for this is that G_v and G_j cross only once at some value in which $R_{jT} = R_{vT}$, corresponding to a stock price denoted by S_T^0, since G_v is clearly above G_j for "small" values of z_0. $H(z_0)$ is estimated by numerical integration, and a non-negative $H(z_0)$ for each z_0 will provide evidence for the improvement in expected utility from adopting the zero-net-cost trading policy for each mispriced option.

Due to the heavy computational requirements, we do not apply in our numerical work the above procedure to all the identified bounds violations in Table 3. Such estimations would have a low informational content, given that they are essentially equivalent to *in-sample* testing. Nonetheless, in *all* the Monte Carlo simulations that were performed we found that $H(z_0)$ had a single zero, thus verifying the conventional SSD criterion (6.1) for the two distributions G_k, $k = v, j$. In fact there are *a priori* reasons to expect the distribution of R_{jT} to exhibit such a single crossing property over that of R_{vT}. We demonstrate this SSD by conditioning on the assignment time τ as well as on the error term ε_τ, and then by comparing the *V*- and *J*-investors' portfolio returns *as if* liquidation has taken place at τ, denoting these returns by $R_{v\tau}$ and $R_{j\tau}$. We also assume that the portfolio revision variables v_t and j_t and their associated transaction costs are approximately equal in the expressions (5.3).[27] These returns now become, in the case of the call option:

[27] Note that Constantinides (1986) has shown that the frequency of trading in the presence of transaction costs is very low, since the investor stays mostly in the NT region.

$$R_{v\tau} = \frac{x_0 R + N_0[S_\tau(1-k_2)+D]}{N_0 S_0(1-k_2+1/\lambda)} \quad ,$$

(6.2)
$$R_{j\tau} = \frac{x_0 R + (N_0+n_0)[S_\tau(1-k_2)+D] - [F_\tau - K]^+}{N_0 S_0(1-k_2+1/\lambda)} \quad ,$$

$$R_E \equiv R_{j\tau} - R_{v\tau} = \frac{n_0[S_\tau(1-k_2)+D] - [F_\tau - K]^+}{N_0 S_0(1-k_2+1/\lambda)} \quad .$$

In (6.2) *D* is the dollar value at τ of dividends *per one share* accruing in a portfolio throughout the life of an option. Replacing F_τ from (2.5) we note that for any given error ε_τ the excess return of the *J*-investor reaches its maximum at $S_\tau = \dfrac{(K-\varepsilon_\tau)}{\alpha_\tau}$ and decreases afterwards. It is positive (negative) to the left (right) of the unique stock price at τ:

(6.3)
$$S_\tau^0 \equiv \{(K-\varepsilon_\tau)/\alpha_\tau + n_0 D\}/[1-(1-k_2)n_0] .$$

It follows, therefore, that for any given value of ε_τ $R_{v\tau}$ and $R_{j\tau}$ cross only once, with $R_{j\tau}$ > (<) $R_{v\tau}$ for low (high) values of S_τ. Further, replacing n_0 from (5.2) in the last expression of (6.2) and noting that the call price C_0 by assumption exceeds the upper bound (2.7)-(2.8), we can easily see that $E[R_{j\tau} | \tau, \varepsilon_\tau] \geq E[R_{v\tau} | \tau, \varepsilon_\tau]$. Hence, all risk averse investors would prefer $R_{j\tau}$ to $R_{v\tau}$. In our empirical work we observe that these properties of $R_{v\tau}$ and $R_{j\tau}$ from the approximate model (6.2) also apply to R_{jT} and R_{vT} given by the exact model (5.3), which considers both transaction costs and any assignment time τ.

We examine several artificial cases of call options under the historical[28] and forward-looking distributions. We compute the upper bound for a call (lower bound for a put) under a given distribution and assume that the observed call price exceeds (lies below) the bound by a factor of 1.25 (0.95), equal to the median observed violation. Then we apply the procedure outlined in the beginning of this section and estimate $H(\infty)$ to measure the utility improvements. For these tests we use as a base case a risk aversion parameter δ = -3 and distributional parameter values typical for our data: the dividend yield γ is 2%, the riskless rate r is 4%.

Figure 4 displays $H(\infty)$, the annualized expected excess return of the *J*- over the *V*-investor as a function of the *portion* of the NT region where the portfolio process has originated separately for calls and puts, with 100% representing the case $y_0 / x_0 = \overline{\lambda}$. Panels A and B plot the expected excess return for all three distributions at- and out-of the money, and panels C and D concentrate on the forward distribution for Sample 2, the one exhibiting the least excess return. Panel C examines five different degrees of moneyness, and panel D varies the risk aversion parameter from δ = -1 to δ = -10.

As expected the results show a strong dependence of the utility improvements on the risk aversion coefficient and, to a lesser degree, on the initial portfolio position. Somewhat less expected is the equally strong dependence on the degree of moneyness of the option and, especially, on the sample used for the estimation of the bounds. Sample 1, which is the low volatility sample, shows the strongest utility improvements. In almost all call option cases, though, the improvements are spectacular for the basic risk aversion parameter of δ = -3, considering the fact that they represent excess returns in a setting in which the annual cum-

[28] Recall that we use separate historical distributions for Sample 1 and 2. Since they are very similar, as may be seen in Table 1, we use only the distribution for Sample 1 in our tests.

dividend risk premium on the index is only 4%. In many instances the expected excess returns of the utility-improving strategies are equal to or larger than 2%, the risky portion of the risk premium on the index. The improvements are much less significant for the put options, for which the median size of the observed violations is significantly smaller.

(Figure 4 about here)

These results are only suggestive as to the validity of our estimated CP (2006) stochastic dominance bounds as utility-improving reservation trading option prices. The utility improvements are dependent on the accurate estimation on the part of the investor of the underlying distribution of the risky asset. The next set of tests, however, admits the imperfect knowledge of the distribution and examines the potential utility improvements from trading on mispriced options *ex post*, given the observed realized return distributions to option expiration, the "best possible information" about the underlying return distribution. This is clearly the equivalent of an out-of-sample test, since these realized distributions were not used in deriving the bounds and identifying their violations.

For these tests we use the violations observed under the two forward-looking distributions. Altogether, we observed 1133 violations for calls, 795 in Sample 1 and 338 for Sample 2; the corresponding numbers for puts were 211 in total, of which all but 13 were in Sample 2. For the two forward-looking distributions we compute the NT region for an investor with a risk aversion coefficient δ=-3 as detailed in section IV, and we consider a starting portfolio at the upper boundary of the NT region, where $y_0 / x_0 = \overline{\lambda}$. For the asset dynamics (2.3') and (2.4), however, we use the realized distribution till the expiration of each option, which is used in Monte Carlo simulations of the path of the ex-dividend return $\frac{S_{t+1}}{S_t}$ till option expiration in order to derive the

$H(\infty)$ utility improvement metric as in (6.1). Appendix D provides an example of this procedure for one particular observed violation.

This test approximates closely the situation facing an investor who computes the option bounds from "noisy" data on the underlying returns and is willing to use them as reservation trading prices to identify mispriced options. The investor's portfolio adjustments take place under the lognormality of returns assumption, since this is the only one for which closed-form expressions exist. In spite of these such an investor would have realized highly significant expected excess returns for call options under the "best possible information set" return distribution: in 871 cases, 76.9% of the total, the expected excess return was positive and the average value of $H(\infty)$ was 1.73%. These numbers reject clearly the null hypothesis that these gains are due to chance with significance levels of less than 1e-5. The results for put options were reversed, with the expected excess return significantly negative, but the numbers were much smaller than for the call options. As a result, the combined sample of violations for both calls and puts gave a proportion of positive expected excess returns of 65.8% and an average expected excess return of 1.05%, both numbers being significant at levels in excess of 1e-5.

Given the importance of this test, we verify its results by applying our trading rule to the *next observed* quote following the violating observation in that same day; we do not verify whether this next quote violates the bound. The total set of observations was reduced to 1087 from 1133, since there were no follow up quotes for some observations within the same day. The results were, not surprisingly, somewhat weakened, but they still remained similar to the ones reported above. For the 925 calls the mean excess return was 1.598%, with a 76.22% probability of a positive excess return for the *J*-investor, highly significant on the basis of the sign test against the null hypothesis of zero excess return. For the 162 puts the results were reversed, but the combined sample of 1087 options again gave highly significant excess returns of 1.206%, with a probability of 68.44% of a positive excess return, again highly significant by the sign test.

An alternative out-of-sample test is the comparison of the policy of writing (purchasing) overvalued call options (undervalued put options) identified through the filters of the option bounds (2.8)-(2.11) against the policy of writing (purchasing) a randomly chosen sample of call (put) options of equal size. For such a test we use the *actual path* of the index futures till option expiration for each one of the mispriced options. We also assume perfect knowledge of the realized futures prices till option expiration in deriving the exercise policy against the *J*-investor, thus biasing the excess return against the *J*-investor.[29] For the 1133 call options violating the upper bounds estimated with the forward distributions this policy yields a total of 693 positive excess returns with an average excess return of -2.14%. In spite of these results a bootstrapped distribution drawn out of our total data shows that there is only a 7.66% probability that a randomly chosen sample of 1133 written calls from our data would outperform these numbers. The repetition of this test with the 925 calls of the next observed quote gave similarly a total of 556 calls with positive excess returns, with an average excess return of – 2.224%, yet the probability of besting these results was only 12.12%. Hence, the call upper bound (2.7)-(2.8) estimated with the forward samples is a valid selector of overvalued call options.

The equivalent results for put options were even more impressive: for the 211 undervalued options the *J*-investor realized positive excess returns in 105 cases with an average excess return of 5.42%. The bootstrapped distribution of excess returns from purchasing 211 put options randomly chosen each time from our data showed that the probability of achieving a better

[29] The tests were also done by constraining assignment or exercise at option expiration. The results were very close to the ones reported here.

performance for the *J*-investor was effectively zero. This result was confirmed with the 162 put options of the next quote sample, implying that the put lower bound (2.10)-(2.11) estimated with the forward samples is a valid filter for identifying undervalued put options.

VII. Conclusions

In this paper we have examined the empirical implications of the Constantinides-Perrakis (2006) futures options bounds in the presence of transaction costs. We used several assumptions about the distribution of the underlying asset, the S&P 500 index. All these assumed distributions were based on histograms drawn from observed market data without imposing any particular class of distributions. We also included the error of the cost-of-carry formula, a key element in the evaluation of the bounds.

First we examined whether observed call (put) futures option bid (ask) prices satisfy the corresponding CP bounds for the two cases where the bounds are tight and relatively invariant to the size of the transaction cost parameters. The bounds were computed on the basis of the estimated index return distribution. We found the best results, the smallest number of violations of the bounds, with a forward looking sample of index values that was broadly time wise coincident with the option sample. Still, a substantial number of violations remained, especially for call options, giving opportunities for realizing stochastically dominant superior returns.

We then examined trading strategies to exploit these observed violations for a representative investor with CPRA utility. For such strategies we measured the improvements in expected utility by constructing a portfolio whose compound return till option expiration would dominate stochastically in the second degree the optimal portfolio of an identical investor who does not adopt the strategy. The compound returns of both investors are net of transaction costs. We find that the expected excess return from the stochastically dominating strategy is significant in all cases, depending on the degree of violation, the underlying distribution, the investor risk aversion and the moneyness of the option. Last but not least, we examined the effects of the adoption of these stochastically dominating strategies for the observed bounds violations in our samples. We found that these strategies would have brought significant excess returns to the CPRA investor even though the underlying asset followed a different distribution than the one assumed in deriving the bounds.

It should be noted that, although the utility improvements were measured under the investor's assumption of a lognormal diffusion asset dynamics, the results are most probably robust against any other investor-hypothesized distribution. As Constantinides (1979) showed, the assumption of a CPRA utility is sufficient to produce a compact no trading region in managing the investor portfolio. For the short-lived options examined in this paper this means that the number of portfolio adjustments and the resulting transaction costs during the life of the option would be very small, thus preserving the static stochastic dominance result described in relations (5.4) and (5.5).

We have, therefore, shown first that the number of observed violations of the CP bounds was relatively small, and second that the stochastically dominant trading strategies when violations are observed do produce significant expected excess returns, both in- and out-of-sample. Furthermore, unlike violations of option pricing results observed under the dominant arbitrage methods, these trading strategies do take place under realistic conditions, with the appropriate transaction costs. Hence, the CP bounds are "legitimate" option bounds in the presence of transaction costs for S&P 500 index futures options.

References

Balduzzi, P., and A. W. Lynch, 1999. "Transaction Costs and Predictability: Some Utility Cost Calculations." *Journal ,of Financial Economics* 52, 47-78.

Bensaid, B., J-P. Lesne, H. Pagés and J. Scheinkman, 1992. "Derivative Asset Pricing with Transaction Costs." *Mathematical Finance* 2, 63-86.

Black, Fischer, 1986. "Presidental Address: Noise." *Journal of Finance* 41, 529-544.

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 Industry 60 Years Later: For Better or Worse?" *Financial Analyst Journal* 61, 15-24.

Boyle, P. P. and T. Vorst, 1992. "Option Replication in Discrete Time with Transaction Costs." *Journal of Finance* 47, 271-293.

Constantinides, G. M., 1979. "Multiperiod Consumption and Investment Behavior with Convex Transactions Costs." *Management Science* 25, 1127-37.

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

Constantinides, G. M., J. C. Jackwerth, and S. Perrakis, 2005. "Mispricing of S&P 500 Index Options," working paper.

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

Constantinides, G. M., and S. Perrakis, 2006. "Stochastic Dominance Bounds on American Option Prices in Markets with Frictions", forthcoming in *The Review of Finance*.

Constantinides, G. M. and T. Zariphopoulou, 1999. "Bounds on Prices of Contingent Claims in an Intertemporal Economy with Proportional Transaction Costs and General Preferences." *Finance and Stochastics* 3, 345-369.

Constantinides, G. M. and T. Zariphopoulou, 2001. "Bounds on Derivative Prices in an Intertemporal Setting with Proportional Transaction Costs and Multiple Securities." *Mathematical Finance* 11, 331-346.

Davis, M. H. A. and A. R. Norman, 1990. "Portfolio Selection with Transaction Costs." *Mathematics of Operations Research* 15, 676-713.

Davis, M. H. A., V. G. Panas and T. Zariphopoulou, 1993. "European Option Pricing with Transaction Costs." *SIAM Journal of Control and Optimization* 31, 470-493.

Dumas, B. and E. Luciano, 1991. "An Exact Solution To A Dynamic Portfolio Choice Problem Under Transactions Costs." *Journal of Finance* 46, 577-596.

Fleming W. L. and R. W. Rishel. *Deterministic and Stochastic Optimal Control* (Berlin, Heidelberg, New York: Springer-Verlag, 1975).

Genotte, G., and A. Jumg, 1994. "Investment Strategies Under Transaction Costs: the Finite Horizon Case." *Management Science* 40, 385-404.

Hadar, J. and W. R. Russell, 1969. "Rule For Ordering Uncertain Prospects." *American Economic Review* 59, 25-34.

Hanoch, G. and H. Levy, 1969. "The Efficiency Analysis Of Choices Involving Risk." *Review of Economic Studies* 36, 335-346.

Hull, J. C. *Options, Futures, & Other Derivatives* (Upper Saddle River: Prentice-Hall, 2003).

Jackwerth, J. C., 2004. *Option-Implied Risk-Neutral Distributions and Risk Aversion,* ISBN 0-943205-66-2, Research Foundation of AIMR, Charlotteville, USA.

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

Kamrad, B., and P. Ritchken, 1991. "Multinomial Approximating Models for Options with k State Variables", *Management Science*, 37, 1640-1652.

Leland, H. E., 1985. "Option Pricing and Replication with Transactions Costs." *Journal of Finance* 40, 1283-1301.

Levy, H., 1985. "Upper and Lower Bounds of Put and Call Option Value: Stochastic Dominance Approach." *Journal of Finance* 40, 1197-1217.

Levy, H., 1992. "Stochastic Dominance end Expected Utility: Survey and Analysis." *Management Science* 38, 555-593.

Liu, H., and M. Loewenstein, 2002. "Optimal Portfolio Selection with Transaction Costs and Finite Horizons." *Review of Financial Studies* 15, 805-835.

Merton, R., (1969), "Lifetime Portfolio Selection Under Uncertainty: the Continuous Time Case", *Review of Economics and Statistics*, 51, 247-257.

Merton, R. C., 1989. "On the Application of the Continuous-Time Theory of Finance to Financial Intermediation and Insurance." *The Geneva Papers on Risk and Insurance* 14, 225-261.

Perrakis, S., 1986. "Option Bounds in Discrete Time: Extensions and the Pricing of the American Put." *Journal of Business* 59, 119-141.

Perrakis, S., 1988. "Preference-free Option Prices when the Stock Return Can Go Up, Go Down, or Stay the Same." *Advances in Futures and Options Research* 3, 209-235.

Perrakis, S., and J. Lefoll, 1997. "Derivative Asset Pricing with Transaction Costs: an Extension." *Computational Economics* 10, 359-376.

Perrakis, S., and P. J. Ryan, 1984. "Option Pricing Bounds in Discrete Time." *Journal of Finance* 39, 519-525.

Ritchken, P. H., 1985. "On Option Pricing Bounds." *Journal of Finance* 40, 1219-1233.

Ritchken, P. H. and S. Kuo, 1988. "Option Bounds with Finite Revision Opportunities." *Journal of Finance* 43, 301-308.

Rothschild, M., and J. E. Stiglitz, 1970. "Increasing Risk: I, a Definition", *Journal of Economic Theory*, 225-243.

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

Silverman, B. W., 1986. *Density Estimation*, Chapman and Hall, London.

Soner, H. M., S. E. Shreve and J. Cvitanic, 1995. "There is no Nontrivial Hedging Portfolio for option Pricing with Transaction Costs. " *Annals of Applied Probability* 5, 327- 355.

Zariphopoulou, T., 1999. "Comment on "The Valuation of Contingent Claims under Portfolio Constraints: Reservation Buying and Selling Prices." *European Finance Review* 3, 389-392.

Appendix

A. Description of the Data

Our primary data set consists of Chicago Mercantile Exchange (CME) tapes containing time-stamped quotes of the S&P 500 futures options and the underlying futures for the period January 1990-December 2002, with the exception of April 1991-December 1991, for which the option data is missing. Some of these options series (quarterly options), namely those that mature the day preceding the maturity of the underlying futures contract[30], have delivery terms that effectively imply that they are exercisable only at the end of the day. Others (serial options), though, mature before the underlying contract and have different delivery terms, which imply that they are exercisable at any time[31]. Accordingly we consider those series of options separately in estimating the bounds. Overall, we sampled quotes at 141 dates for which 20825 (27085) raw call (put) quotes were recorded. Since the prescribed investment policies require the investor to sell (purchase) calls (puts), we collected bids (asks) for calls (puts). CME, however, flags bids (asks) only if a bid (ask) is higher (lower) than the preceding bid (ask); in addition, no transaction data is flagged. These data characteristics would result in a number of available call (put) quotes of 3450 (5141). To augment the data set, we recovered flags from non-transacted data for bids (asks) that were lower (higher) than the preceding non-transacted

[30] S&P 500 futures options mature at the futures trading close on Thursday preceding the third Friday of a month.

[31] The S&P futures contracts mature in the March quarterly cycle, which implies that there is twice as much sampling dates for serial than for quarterly options.

quote in the daily series for a given strike[32]. The original and recovered quotes entered our sample provided the futures quote could be matched within 10 seconds. As the final screen, we entered in the sample only those options within the moneyness range of 0.9-1.05, with the moneyness defined as the F/K (K/F) ratio for calls (puts). This resulted in the final sample of 7001 calls and 8310 puts.

To derive the option bounds, we originate the index process from the index value derived from the futures price contemporaneous to an option quote under the perfect cost-of-carry relationship. It has been argued (see, for instance, Jackwerth and Rubinstein, (1996)) that the futures quotes provide a better proxy for the S&P then the spot index since futures are a traded asset, and it was shown that the spot index quotes lag the index values. In turn, we use this futures implied index value to derive the value of the underlying futures contract under the cost-of-carry model (2.6)-(2.7), which assumes the existence of an error term in the relationship. We select an empirical proxy for this error term. Tables A1 and A2 display the distribution of the error from the cost-of-carry relationship for 1990-2002, respectively for intraday observations and for the futures settlement price. Quotes used to derive the departures from the exact cost-of-carry relationship were sampled eight times a day at fixed hours for intraday errors by finding the best time match between the cash and futures for a 30-second interval around the sampling time[33]. Cash index quotes were obtained from CME. From these tables it is clear that an error $\bar{\varepsilon}$ = 1% from the perfect cost-of-carry futures price would include all but a few outliers of the observed error during the entire period. A value of $\bar{\varepsilon}$ = 0.5% would lie beyond three standard

[32] As a safeguard against misreading flags, we verified that the results for the recovered were not qualitatively different from those for the exchange flagged quotes.

[33] The length of this interval was set to guarantee finding quotes for both assets.

deviations above the mean and the median for the settlement price and would include more than 99% of the observed errors. For intraday observations such a value would lie at above two standard deviations from both mean and median and would include around 95% of the observed errors. We use this last value of $\bar{\varepsilon} = 0.5\%$ of the futures price to proxy for ε_t in (2.6) to transform the modeled index process into the value of the underlying futures at each time t in (2.8)-(2.9) and (2.10)-(2.11). We also assume that this empirical proxy $\bar{\varepsilon}$ will also compensate for imperfect simultaneity in index and futures quotes, as well as for the bid-ask spread in futures at the matching of an option and a futures quote. Notice that it is apparent from (2.8)-(2.9) and (2.10)-(2.11) that a positive error term increases (decreases) the upper (lower) bound.

(Tables A1 and A2 about here)

B. <u>A Lattice Model for the Estimation of the Bounds</u>

To derive the bounds (2.8)-(2.11) we use empirical distributions of daily S&P 500 returns estimated from the samples described in section III. Since the estimation of the bounds becomes very quickly explosive with any realistic number of distinct states in the daily returns distribution, we use a lattice method to aggregate numerically similar index returns. The procedure is summarized below, with a full description of the method available from the authors on request.

Consider a discrete distribution with n states in one period. The convolution of this distribution with itself produces n^m returns after m time periods. We begin by reducing this number of returns by recognizing partial recombinations of the lattice that produce identical returns. The probabilities of these returns are aggregated at each time step, thus reducing the size of the tree at each period by dealing only with unique returns. Still, the size of the tree remains explosive for the values of distinct states n used in this paper.

The next step in simplifying the computations is to aggregate together into a single state numerical returns that are distinct but "close" to each other. Suppose that at a given time step there are N distinct states, whose values and probabilities are denoted by the pair (r, p) of vectors in $\Re_N^+$. We reduce these vectors into M-dimensional vectors (R, q) in $\Re_M^+$, M < N, by defining

$$R_j = q_j^{-1} \sum_{k=1}^{m_j} p_k r_k ;\ q_j = \sum_{k=1}^{m_j} p_k ,\ j = 1 \ldots M, \tag{B.1}$$

where (r_k, p_k) satisfy the relation $(K - 0.5)\varepsilon \le r_i < (K + 0.5)\varepsilon$ for K = 0, 1,…Ň, with Ň being the smallest number sufficient to cover the largest of the N returns, and $\varepsilon \in \mathbb{R}^+$ is a rounding factor.

It can be shown that that for a suitable rounding factor ε, the reduced distribution $(R; q)$ defined in (B.1) approximates well[34] the full distribution in the context of derivatives pricing. Clearly, the accuracy of the approximation, but also the computational time and difficulty, vary inversely with the size of ε. Fortunately, for our data and for the time horizon chosen in the estimation of the bounds a value of ε may be found such that the reduction in the lattice size yields both computational precision and relative speed.

In the numerical estimation of the bounds we use a 50-state daily returns distribution, which must be convoluted approximately 21 times with itself in order to produce bounds for 30-day options according to (2.8)-(2.11). The full terminal distribution contains 1.62e17 unique terminal returns, which are clearly beyond computational capacity. On the other hand, a value of ε = 1e-4 in (B.1) reduces the terminal states to about 3e4, which allow both feasible and accurate

---

[34] The rounding scheme in (B.1) yields the same first moment and decreases only slightly the magnitudes of the second, third and fourth moments of the full distribution, even though a reduction occurs at each time step.

computations of the bounds.[35] We present an example of the accuracy of the approximation in Table B1, which shows the computation of a stylized call upper bound (2.8)-(2.9) evaluated with the empirical distribution extracted from the 1997-2002 S&P 500 daily returns.

(Table B1 about here)

In the table we use six different values of ε, varying from 1e-2 to 1e-5, and three different daily returns distributions, with 8, 21 and 50 states respectively; the 8-state distribution allows exact results in the computation of the bound. It is clear that for the four smallest sizes of ε shown in the table, from 1e-3 to 1e-5, the estimated call option bounds are virtually indistinguishable from each other in all cases. Observe also that the bounds increase monotonically as ε decreases, implying that the bounds converge uniformly from below to their "true" value. For the chosen value of ε = 1e-4 that we used in our numerical work the increase in the size of the bound as we move to the next lowest value of 1e-5 is virtually non-existent.[36] This is all the more surprising given that the reduction in the number of terminal nodes achieved by applying (B.1) is impressive: Panel B of Table B1 shows the reduced number of nodes as a proportion of the original number of nodes, amounting for the case of interest to a miniscule proportion of 7.50e-14.

[35] The detailed numerical algorithm is available from the authors on request.

[36] The convergence of the bound computed under the transformation (B.1) to its "true" value was also verified by Monte Carlo simulations of *European* options, by setting the dividend rate equal to zero in (2.8)-(2.9). The Monte Carlo method turned out to be less accurate than the transformation (B.1).

C. The Constantinides (1986) Model with Dividends

In this appendix we extend the Constantinides (1986) model to incorporate constant yield cash dividends on the risky asset. Assuming continuous deterministic dividend yield at the rate γ, the assets' dynamics now become, instead of (2.1)

$$
\text{(C.1)} \qquad \begin{aligned} dx &= (r-\beta)xdt + \gamma ydt \\ dy &= \mu ydt + \sigma yd\omega \ . \end{aligned}
$$

For (C.1), the resulting Bellman equation has no known closed-form solution[37]. To estimate the adjusted value function $V(x_t,\ y_t,\ t)$ in the case of dividends, we approximate the bond account dynamics by substituting in the first equation of (C.1) instead of y the value of x multiplied by λ^*, the optimal y to x proportion inclusive of dividends without transaction costs or *Merton line* after Merton (1969):

$$
\text{(C.2)} \qquad \lambda^* = \left[\frac{\mu+\gamma-r}{(1-\delta)\sigma^2}\right]\left[1-\frac{\mu+\gamma-r}{(1-\delta)\sigma^2}\right]^{-1}.
$$

The approximate bond dynamics are:

$$
\text{(C.3)} \qquad dx = (r-\beta+h)xdt\,,
$$

where $h \equiv \gamma\lambda^*$. This gives us the system (4.2).

To prove now the equations for the value function (4.5)-(4.7) we note that under (C.1)-(C.3) the Bellman equation by Ito's lemma becomes:

$$
\text{(C.4)} \quad (\beta x)^{\delta}/\delta + (r-\beta+h)xV_x^{\gamma} + \mu yV_y^{\gamma} + (\sigma^2/2)y^2V_{yy}^{\gamma} - \rho V^{\gamma} = 0, \quad \underline{\lambda} \le y/x \le \overline{\lambda}\,.
$$

[37] A review of the dynamic programming literature has revealed that a closed form solution to the Bellman equation rarely exists. See Fleming and Rishel (1975).

If we substitute the value function (4.4) and its appropriate partial derivatives into (C.4) and simplify, we obtain $A_1 (y/x)^{s_1} f(s_1) + A_2 (y/x)^{s_2} f(s_2) = 0$, where f(.) is as defined in (4.5). Since for (D.4) to hold we must have $f(s_1) = f(s_2) = 0$, the equation (4.5) follows immediately. Substituting the value function V^{γ} into the boundary conditions (4.6ab) and dividing by x^{δ} yields the same pair of equations (4.7ab) for (A_1, A_2) as in the no-dividend case.

D. <u>An Illustration of an Out-of-Sample test</u>

On August 8th, 1990 we observed a call bid price of 10.10. The estimated call upper bound with the corresponding forward-looking sample was 8.98. The contemporaneous futures price was 334.45 and the futures implied index was 333.41. The annualized dividend yield till the expiration of the option and the observed riskless rate were, respectively, 2.85% and 7.13%. Hence, for a total excess return of 4% over the riskless rate the corresponding ex-dividend risk premium would be approximately 1.15%. Last, the estimated boundaries of the NT region were 1.22 and 9.58.

For that particular option we observed 22 daily index returns till expiration. After subtracting the mean and adding the riskless rate and the 4% *cum*-dividend premium we find the following ordered series of returns:

0.97520, 0.98031, 0.98552, 0.98576, 0.98649, 0.98772, 0.98852, 0.98984, 0.99155, 0.99565, 0.99813, 1.00140, 1.00142, 1.00279, 1.00608, 1.00829, 1.01260, 1.01398, 1.01740, 1.01881, 1.02530, 1.03098

From this series, by sampling 22 times with replacement we generate a sequence of daily returns representing a possible path under the realized returns distribution; we include intraday trading in this path by the procedure described in Section III. We then apply (5.3) to this path and estimate the returns R_{jT} and R_{vT} of the *J*- and *V*-investors, starting the portfolios at the upper boundary $\overline{\lambda}$ of the NT region with 1.5 shares for one written call in order to satisfy the monotonicity condition. We repeat the sampling and estimation 1e5 times in order to construct the distributions of R_{jT} and R_{vT} under the "best possible information" index return distribution for that particular observation. After verifying the single crossing of the two distributions, we estimate the expected excess return to the *J*-investor, which turned out to be 3.27% for that observation. Note that for the return to the *J*-investor we also need to stipulate the exercise policy, which is given by the function $N(S_t, t)$ in (2.8); this function needs therefore to be estimated separately, and its comparison to the immediate exercise value yields the exercise policy along each path.

Table 1. Second, Third and Fourth Moments of S&P 500 Daily Returns

Panel A: Historical Daily Returns				
	Sample 1		Sample 2	
	Observed (N=16633)	Aggregated (N=150)	Observed (N=18704)	Aggregated (N=150)
Volatility	0.18474	0.18467	0.17921	0.17914
Skewness	-0.47747	-0.47796	-0.48068	-0.48123
Kurtosis	22.36602	22.33360	22.94456	22.97990
Panel B: Forward-looking Daily Returns				
	Sample 1		Sample 2	
	Observed (N=2071)	Aggregated (N=50)	Observed (N=1509)	Aggregated (N=100)
Volatility	0.11491	0.11486	0.21223	0.21219
Skewness	-0.17328	-0.17354	-0.06498	-0.06499
Kurtosis	2.29656	2.30192	2.12150	2.12329

Table2. Summary Statistics for S&P 500 Daily Returns Realized until Option Maturity

	Mean	Median	Min	Max	St. Dev.
	Panel A: Sample 1 (N=70)				
Volatility	0.1046	0.1011	0.0499	0.2418	0.0374
Skewness	-0.0099	0.0761	-1.7197	1.2739	0.5509
Kurtosis	0.2149	-0.0788	-1.2617	4.0116	1.1221
	Panel B: Sample 2 (N=71)				
Volatility	0.1963	0.1825	0.0882	0.4439	0.0719
Skewness	0.0570	0.0748	-1.0401	0.9904	0.4872
Kurtosis	-0.1194	-0.2995	-1.1988	3.2578	0.8391

Table 3. Violations of the Bounds (2.8)-(2.9) and (2.10)-(2.11)

The table displays the proportions of the option quotes violating the stochastic dominance bounds (2.8)-(2.9) and (2.10)-(2.11). The top entry corresponds to the bounds estimated by the distribution of the S&P 500 daily returns; the bottom entry corresponds to the bounds estimated by the lognormal distribution with the same volatility. The entries in brackets represent the percentage of sampling dates in each sample for which violations of the bounds were found.

Index Distribution	Calls: Pr. Violations (%) (Risk Premium = 4%)	Calls: Pr. Violations (%) (Risk Premium = 8%)	Puts: Pr. Violations (%) (Risk Premium = 4%)	Puts: Pr. Violations (%) (Risk Premium = 8%)
Panel A: Sample 1990 – 1996				
	(N=3826)		(N=5620)	
Historical	5.3 (6) 5.3 (6)	4.4 (6) 4.2 (6)	63.3 (77) 65.0 (83)	52.3 (71) 56.8 (74)
Forward	20.8 (30) 21.7(31)	16.0 (23) 16.2 (24)	0.2 (1) 0.2 (1)	0.0 (0) 0.0 (0)
Realized	34.3 (54) 37.3 (61)	20.6 (43) 21.7 (44)	0 (0) 0 (0)	0.0 (0) 0.0 (0)
Panel B: Sample 1997 – 2002				
	(N=3175)		(N=2690)	
Historical	23.4 (53) 24.6 (41)	22.8 (41) 18.8 (36)	0.0 (1) 0.1 (3)	0.0 (0) 0.0 (0)
Forward	8.8 (26) 10.7 (26)	5.9 (17) 6.4 (17)	7.4 (21) 7.7 (20)	2.7 (11) 3.1 (13)
Realized	35.3 (47) 35.9 (50)	23.5 (37) 26.6 (39)	12.1 (14) 12.8 (16)	11.0 (11) 10.9 (13)
Panel C: Sample 1990 – 2002				
	(N=7001)		(N=8310)	
Historical	13.5 (29) 14.1 (24)	12.7 (24) 10.8 (21)	43.3 (39) 46.5 (43)	35.4 (35) 38.4 (37)
Forward	15.3 (28) 16.7 (29)	11.4 (20) 11.7 (21)	3.4 (11) 5.2 (11)	0.9 (6) 1.0 (6)
Realized	34.7 (51) 36.7 (56)	21.9 (40) 23.9 (41)	5.1 (7) 5.4 (8)	3.6 (6) 3.5 (6)

Table 4. No Transaction Region

The table displays the proportions of the risky to riskless asset within which it is optimal to refrain from trading derived by the Constantinides (1986) model with dividends. The dividend yield γ is 2%, the riskless rate r is 4%. The second moments are as presented in Table 1.

Distribution	δ	$\underline{\lambda}$	$\bar{\lambda}$	Merton Line
Historical	-3	0.306	0.418	0.415
	-5	0.177	0.244	0.243
	-10	0.087	0.120	0.119
Forward 1	-3	0.999	6.013	3.150
	-5	0.885	1.041	1.024
	-10	0.319	0.383	0.381
Forward 2	-3	0.200	0.288	0.286
	-5	0.121	0.175	0.174
	-10	0.061	0.089	0.088

Table A1. Descriptive Statistics of Intraday Error from Cost of Carry for 1990-1996

Statistics refer to the error ε from the cost of carry relationship $F = S \exp[(r-\gamma)T + \varepsilon]$, where F is the price of futures underlying serial options, S is the cash index quote contemporaneous with the futures quote, r is the observed 3-month T-bill rate, γ is dividend yield based on daily dividends until the futures expiration date, T is the time till the futures expiration. The error was sampled in hour-long intervals starting at the open (8:30AM), as well as at the close (3:15PM) for the total of eight times for each day before the expiration of the serial options in the sample. Quotes used in the error estimation were found by the best time match between the cash index and futures quotes within a 30-second interval around the sampling time.

N	15026	Max	0.0247
Mean	1.78E-03	99%	0.0081
Std Dev.	2.24E-03	95%	0.0052
Skewness	-0.5516	90%	0.0041
Kurtosis	21.6419	Q3	0.0028
Std Err.	1.83E-05	Median	0.0016
Pr(Mean=0)	<0.0001	Q1	0.0007
Range	0.0621	10%	-0.0001
Q3-Q1	0.0021	5%	-0.0009
		1%	-0.0041
		Min	-0.0374

Table A2. Descriptive Statistics of Error from Cost of Carry at the Futures Settlement for 1997-2002

Statistics refer to the error ε from the cost of carry relationship $F = S\exp[(r\text{-}\gamma)T + \varepsilon]$, where F is the price of futures underlying quarterly options, S is the cash index quote contemporaneous with the futures quote, r is the observed 3-month T-bill rate, γ is dividend yield based on daily dividends until the futures expiration date, T is the time till the futures expiration. Aggregated cash index quotes for a 30-second interval before the close (3:15PM) were compared to the futures settlement price for each day before the expiration of the quarterly options in the sample.

N	953	Max	0.0066
Mean	2.77E-04	99%	0.0045
Std Dev.	2.14E-03	95%	0.0030
Skewness	-3.7014	90%	0.0024
Kurtosis	45.1780003	Q3	0.0014
Std Err.	6.92E-05	Median	0.0004
Pr(Mean=0)	<0.0001	Q1	-0.0006
Range	0.0369	10%	-0.0019
Q3-Q1	0.0021	5%	-0.0029
		1%	-0.0061
		Min	-0.0302

Table B1. Performance of Reduced Lattice Model

The table displays the performance of the reduced discrete distributions (B.1). The starting index value is 1000, strike is 1000, riskless rate is 4%, risk premium is 4%, and dividend yield is 2%. The maturity of 30 days is partitioned into 21 trading periods.

ε	n = 8	n = 21	n = 50
Panel A: Call Bound Values			
1e-2	25.63721	26.48949	26.54064
2e-3	25.64199	26.84576	27.00539
1e-3	25.64210	26.85478	27.01661
2e-4	25.64227	26.85500	27.01687
1e-4	25.64227	26.85504	27.01691
1e-5	25.64227	26.85504	27.01692
0	25.64227	N.A.	N.A.
Panel B: Pr. of Exact Distribution Terminal Nodes			
1e-2	2.00e-04	1.04e-09	7.63e-16
2e-3	9.49e-04	5.21e-09	3.84e-15
1e-3	1.83e-03	1.03e-08	7.64e-15
2e-4	8.03e-03	4.97e-08	3.78e-14
1e-4	1.50e-02	9.76e-08	7.50e-14
1e-5	1.15e-01	9.17e-07	7.30e-13

Figure 1. Kernel Regression of the Implied Volatility of Observed Option Prices

The figure displays implied volatility for call bidding prices (put asking prices) for Samples 1 and 2 respectively in series Bid 1 and Bid 2 (Ask 1 and Ask 2).

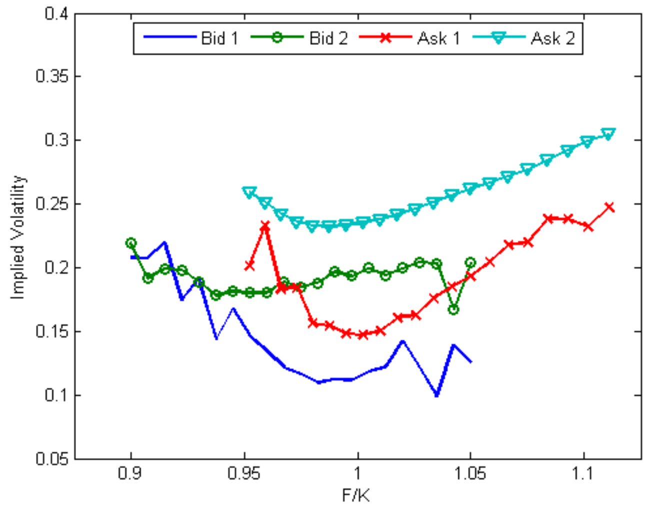


Figure 2. Kernel Regressions of the Call and Put Bounds together with the Observed Option Quotes

The figure displays kernel regression (solid lines) of the implied volatility of the bounds (2.8)-(2.9) and (2.10)-(2.11). Index distributions applied to derive the bounds are as described in Section III. Dots represent the implied volatility of observed option quotes.

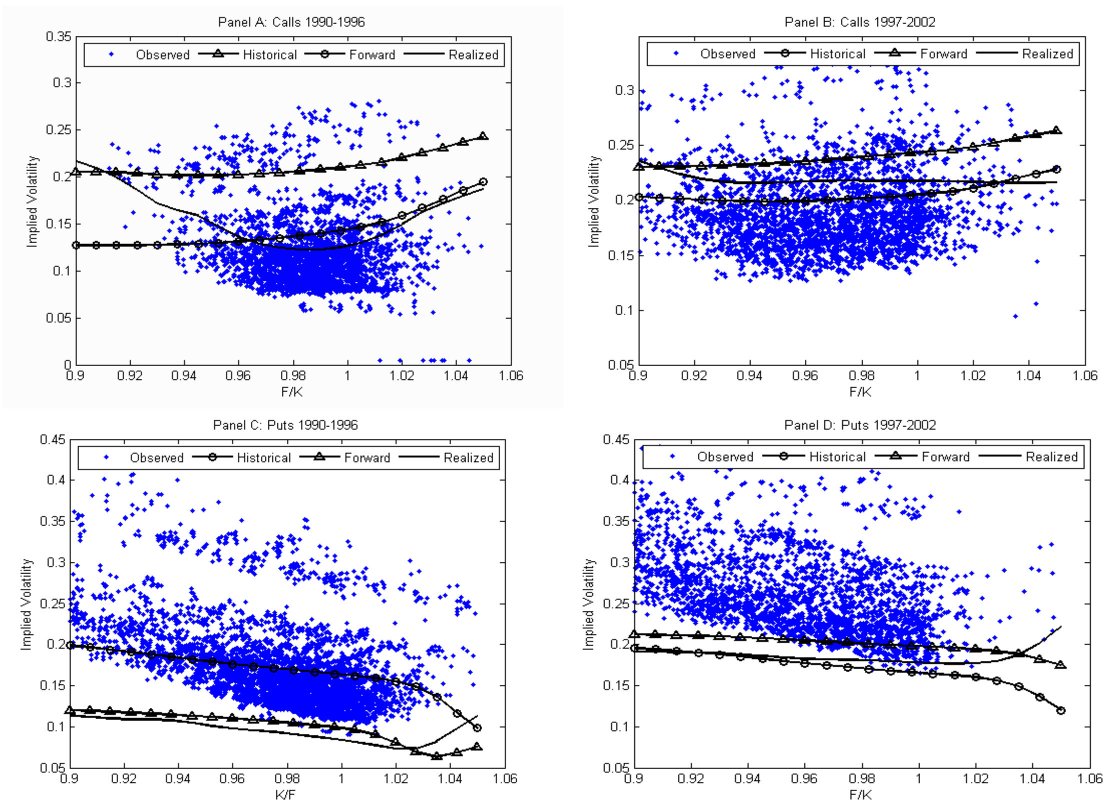
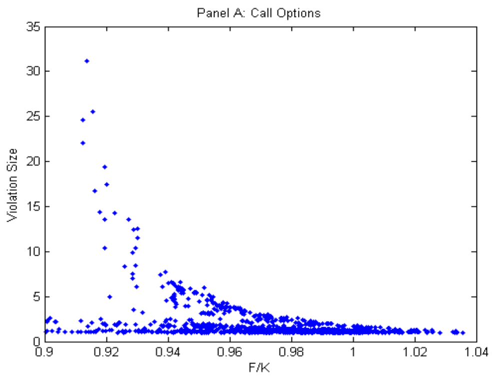


Figure 3. Moneyness and Bounds Violations

The figure displays the ratio (Violation Size) of the observed option quotes to the bounds (2.8)-(2.9) and (2.10)-(2.101 in relation to moneyness. Results are for those quotes that were mispriced under forward-looking distributions as described in Section III.



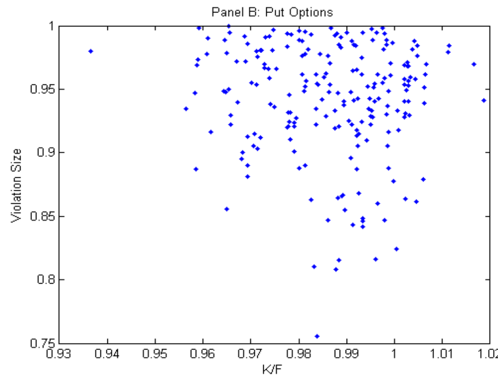


Figure 4a. Trading on Mispriced Call Options

The figure displays the expected utility improvement metric $H(\infty)$ (6.1) for trading in mispriced index futures call options under the assumption that the distribution used to derive the call upper bound (2.8)-(2.9) generates the index return till the option expiry. Index distributions applied to derive the bounds are as described in Section III. The ratio of a call price to its upper bound is set to 1.25. Other data is as follows: riskless rate 4%, risk premium 4% and dividend yield 2%.

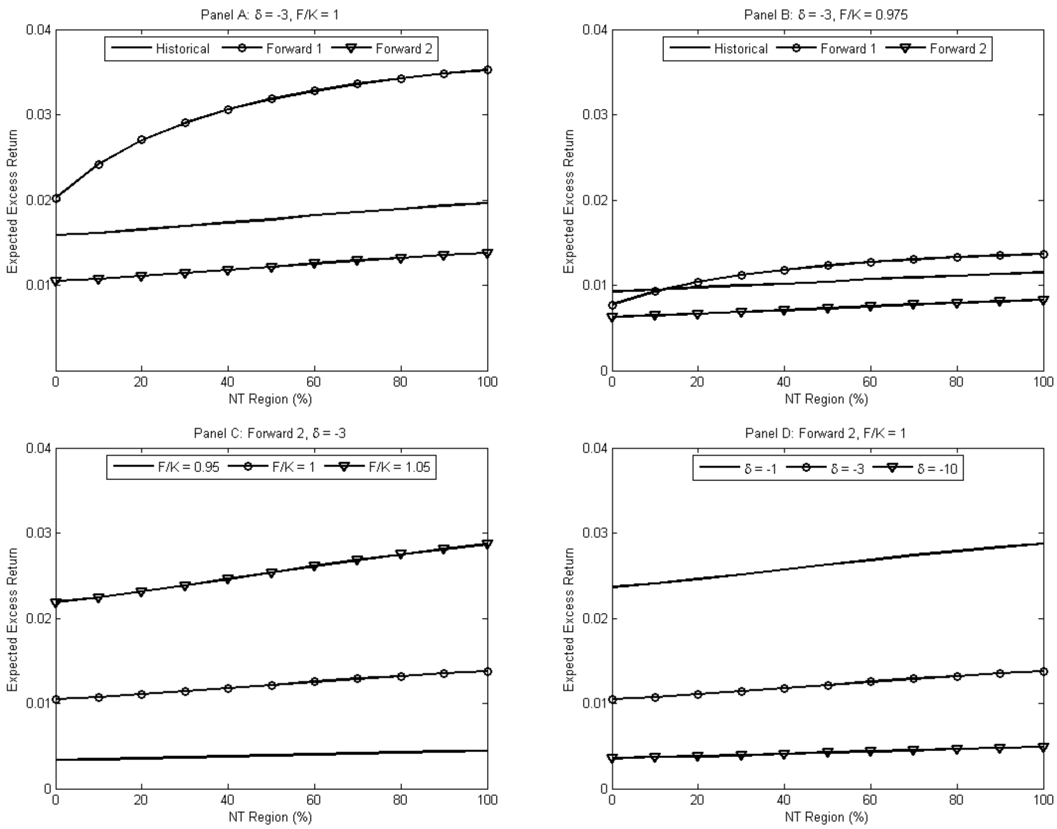


Figure 4b. Trading on Mispriced Put Options

The figure displays the expected utility improvement metric $H(\infty)$ (6.1) for trading in mispriced index futures put options under the assumption that the distribution used to derive the put lower bound (2.10)-(2.11) generates the index return till the option expiry Index distributions applied to derive the bounds are as described in Section III. The ratio of a put price to its lower bound is set to 0.95. Other data is as follows: riskless rate 4%, risk premium 4% and dividend yield 2%.