Why are Put Options So Expensive?*

Oleg Bondarenko†

Abstract

This paper studies the "overpriced puts puzzle" – the finding that historical prices of the S&P 500 put options have been too high and incompatible with the canonical asset-pricing models, such as CAPM and Rubinstein (1976) model. Simple trading strategies that involve selling at-the-money and out-of-the-money puts would have earned extraordinary profits. To investigate whether put returns could be rationalized by another, possibly nonstandard equilibrium model, we implement a new methodology. The methodology is "model-free" in the sense that it requires no parametric assumptions on investors' preferences. Furthermore, the methodology can be applied even when the sample is affected by certain selection biases (such as the Peso problem) and when investors' beliefs are incorrect.

We find that *no* model within a fairly broad class of models can possibly explain the put anomaly.

JEL Classification: G12, G13, G14

Keywords: Market Efficiency Hypothesis, Rational Learning, Option Valuation, Risk-Neutral Density, Peso Problem

---

*Some results of this paper first appeared under title "On Market Efficiency and Joint Hypothesis." I would like to thank for helpful comments Peter Bossaerts, George Constantinides, Bing Han, Dietmar Leison, Jun Pan, Bill Zame. I am also grateful for the input of seminar participants at Caltech, INSEAD, McGill University, Pennsylvania State University, Stockholm School of Economics, Vanderbilt University, University of Illinois at Chicago, University of Texas at Dallas, University of Wisconsin at Madison, Yale University, the CEPR European Summer Symposium in Financial Markets 2001, American Finance Association meetings 2004, and the conference on "People and money: The human factor in financial decision-making" in Chicago. Any remaining errors are mine.

†Department of Finance (MC 168), University of Illinois at Chicago, 601 S. Morgan St., Chicago, IL 60607. Phone: (312) 996-2362. Email: olegb@uic.edu

1 Introduction

Historical returns of the US equity put options are puzzling. Over the period from 08/1987 to 12/2000, put options on the S&P 500 futures appear to be grossly overpriced. For example, Table 1 reports that puts with one month to maturity have highly negative and statistically significant excess returns. The average excess return is -39% per month for at-the-money (ATM) puts and is -95% per month for deep out-of-the-money (OTM) puts. This implies that selling unhedged puts would have resulted in extraordinary paper profits over the sample period. Other striking findings about historical put returns are that:

- The Jensen's alpha for ATM puts is -23% per month and highly significant. Other popular measures like the Sharpe ratio, the Treynor's measure, the M-squared measure also indicate that put prices have been very high.
- For ATM puts to break even (i.e., to have the average excess return of zero), crashes of the magnitude experienced in October 1987 would have to occur 1.3 times per year.
- The economic impact of the put mispricing appears to be substantial. We estimate the cumulative wealth transfer from buyers to sellers of the S&P 500 futures options and find it to be astounding $18 bln over the studied period.

There is no arguing that selling naked puts could be very risky. For example, a short position in ATM put has a highly asymmetric payoff profile, with limited upside and essentially unlimited downside. Such a position makes a small profit most of the time, but takes a big loss once in a while. Furthermore, the position makes money in good states of the world and loses money in bad states. Because puts are negatively correlated with the market, it is not surprising that they are traded at *negative* risk premiums. Moreover, because of considerable leverage, the magnitude of those risk-premiums is expected to be large.

While it is clear that option traders will only sell puts when properly rewarded for bearing substantial risks, it is much less clear what their normal risk compensation should be. Stated differently, is about 40% per month represents a "fair" return for a short position in ATM puts? Or, perhaps, it is too high. The answer to this question depends on the assumed equilibrium model, as different models predict different risk premiums. In this paper, we initially consider two candidate asset pricing models – CAPM and Rubinstein (1976) – and argue that historical put prices are far too high to be compatible with those canonical models. This does not immediately mean that option markets are irrational, for it is possible that there is another, *nonstandard* equilibrium model which could rationalize the empirical findings.

We explore three natural explanations for the "overpriced puts" anomaly:

E1: *Risk premium.* According to this explanation, high prices of puts are *expected* and reflect normal risk premiums under some equilibrium model. Even though the standard models cannot explain the data, maybe there is another model which can. In this "true" model, investors strongly dislike negative returns of the S&P 500 Index and are willing to pay hefty premiums for portfolio insurance offered by puts.

E2: *The Peso problem.* According to this explanation, the sample under investigation is affected by the Peso problem. The Peso problem refers to a situation when a rare but influential event could have reasonably happened but did not happen in the sample.

To illustrate this explanation, suppose that market crashes (similar to that of October 1987) occur on average once in 5 years. Suppose also that investors correctly incorporate a probability of another crash in option prices. However, since only one market crash has actually happened over the studied 14-year period, the *ex post* realized returns of the Index are different from investors' *ex ante* beliefs. In this case, puts only *appear* overpriced. The mispricing would have disappeared if data for a much longer period were available.

E3: *Biased beliefs.* According to this explanation, investors' subjective beliefs are mistaken. Similar to E2, this explanation states that the Index realized returns have not been anticipated by investors.

Consider an example. Suppose that the *true* probability of a crash in a given year is 20%, but investors incorrectly believe that this probability is 40%. Since investors overstate probabilities of negative returns, puts (especially OTM puts) are too expensive.

To test whether explanations E1-E3 have merit, we implement a new methodology proposed in Bondarenko (2003a). The methodology can best be explained on a simple example.

Consider a finite-horizon, pure-exchange economy with a single risky asset, traded in a frictionless market on dates $t = 0, 1, \ldots, T$. The asset's price is v_t, and the risk-free rate is zero. There exists a representative investor who maximizes the expected value of the utility function $U(v_T)$. Let Z_t denote the value of a general derivative security with a single payoff Z_T at time-T. The security's price satisfies the standard restriction:

$$E_t[Z_s m_s] = Z_t m_t, \qquad t < s, \tag{1}$$

where $m_t = E_t[m_T]$ is the pricing kernel, and $E_t[\cdot]$ is the objective expectation. Traditional tests of the Efficient Market Hypothesis (EMH) are based on the restriction in (1). In those tests, one must pre-commit to a specific pricing kernel, which is usually obtained from a parametric equilibrium model. As a result, tests suffer from a joint hypothesis problem: rejections may emerge because the market is truly inefficient or because the assumed model is incorrect.

Bondarenko (2003a) shows that, under fairly general conditions, securities prices must satisfy another martingale restriction. Let $h_t(v_T)$ denote the conditional risk-neutral density (RND) of the asset's final price. Then securities prices deflated by RND evaluated at the final price are martingales:

$$E_t^v\left[\frac{Z_s}{h_s(v)}\right] = \frac{Z_t}{h_t(v)}, \qquad t < s < T, \tag{2}$$

where $E_t^v[\cdot] := E_t[\cdot\,|\tilde{v}_T = v]$ denotes the expectation *conditional* on the final price being v. Intuitively, the restriction in (2) says the following. Suppose that the empiricist observes many repetitions of the same environment and selects only price histories for which $\tilde{v}_T = v$. Then, in those histories the ratio $Z_t/h_t(v)$ must change over time unpredictably.[1]

Note that the restriction assumes that time-series of $h_t(v_T)$ is available to the empiricist. Despite the fact that RND is not directly observable in financial markets, it is implicit in securities prices. In particular, RND can be estimated from prices of traded options, such as standard calls with different strikes.

[1] In an important paper, Bossaerts (2003) demonstrates that conditioning on future price outcomes can be useful in testing asset pricing models. Bondarenko (2003a) generalizes his results from risk-neutrality to general risk preferences. Bondarenko (1997), Bondarenko and Bossaerts (2000), Bossaerts (2003, 1999), Bossaerts and Hillion (2001) present empirical applications based on the theory in Bossaerts (2003).

The new restriction in (2) has three unique properties. First and most significantly, the restriction makes no reference to the pricing kernel. In other words, it is preference-independent: the utility function $U(v_T)$ can be arbitrary and the restriction in (2) must still hold. This implies that the restriction in (2) can be used to resolve the joint hypothesis problem present in EMH tests. It allows one to test whether securities prices are compatible with *any* equilibrium model from a broad class of models.

Second, the restriction in (2) can be used in samples which come with various selection biases. To see this more clearly, suppose that the empiricist has collected a dataset in which not all price histories are present. For example, suppose that the dataset includes only those histories for which the asset's final price is greater than the initial price, $v_T \geq v_0$. Such a deliberate selection bias will normally cause rejection of (1), even if the true pricing kernel m_t were known. Interestingly, the selection bias does *not* affect the restriction in (2). This is because the restriction involves conditioning on the final price. By the same reason, the restriction is also not affected by the Peso problem discussed in E2.

Third, the restriction in (2) continues to hold even when investors' beliefs are mistaken. Specifically, suppose that investors have incorrect expectations about the distribution of v_T but they update their expectations in a rational way.[2] Then, under a certain additional condition, the restriction in (2) must still hold.

To summarize, there are two alternative approaches for testing rationality of asset pricing. The first one is based on the standard restriction in (1). In this approach, the empiricist must know the true preferences of investors. The approach works *only* if investors' beliefs are correct and the sample is unbiased. The second approach is based on the new restriction in (2). In this approach, the empiricist does need to specify investors' preferences. Investors' beliefs may be mistaken. In fact, preferences and beliefs may even change from one history to another. Moreover, the empiricist may use samples affected by the Peso problem and some other selection biases.

By following the second approach, we are able to verify whether explanations E1-E3 can account for high prices of the S&P 500 puts. If investors are rational and put returns are low because of some combination of E1-E3, then the restriction in (2) must hold.

Empirically, however, the new restriction is strongly rejected. This means that *no* equilibrium model from a *class* of models can possibly explain the put anomaly, even when allowing for the possibility of incorrect beliefs and a biased sample. The class of rejected models is fairly broad. In particular, it includes equilibrium models for which the pricing kernel $m_T = m(v_T)$ is a flexible and unspecified function of v_T. This is an important benchmark case in the theoretical literature. More generally, rejected pricing kernels can also depend on other state variables besides v_T, provided that *projections* of the kernels onto v_T are *path-independent*. Our empirical findings have important implications for the option pricing literature, in particular, for the literature on recovering implied risk preferences from option prices.

The remainder of the paper is organized as follows. Section 2 describes the dataset of S&P 500 futures options and documents the overpriced put puzzle. Section 3 first reviews and extends the theory developed in Bondarenko (2003a), and then implements the model-free approach based on the new restriction in (2). Section 4 discusses the implications of the empirical results and Section 5 concludes.

[2]The extension of EMH where investors are rational but may have incorrect beliefs is studied in Bossaerts (2003, 1999). The extension is termed *Efficiently Learning Market* (ELM).

2 Historical Option Returns

This section documents the overpriced puts puzzle. We start by establishing the framework. Then, we discuss the data and report the empirical results.

2.1 Preliminaries

Let v_t denote time-t value of the S&P 500 Index. We will study historical returns of options written on the Index. Therefore, let $P(K) = P(K,T;v_t,t)$ and $C(K) = C(K,T;v_t,t)$ be the prices of European put and call with strike K and maturity T. To simplify exposition, we assume throughout the paper that the risk-free rate is zero and that S&P 500 does not pay dividends.[3] The option prices can be computed using the risk-neutral density (RND):

$$P(K) = \int_0^\infty (K - v_T)^+ \, h(v_T)\, dv_T, \qquad C(K) = \int_0^\infty (v_T - K)^+ \, h(v_T)\, dv_T,$$

where $h(v_T) = h(v_T, T; v_t, t)$ is RND. RND satisfies the relationship first discovered in Ross (1976), Breeden and Litzenberger (1978), Banz and Miller (1978):

$$h(v_T) = \left. \frac{\partial^2 P(K)}{\partial K^2} \right|_{K=v_T.} = \left. \frac{\partial^2 C(K)}{\partial K^2} \right|_{K=v_T.} \tag{3}$$

This relationship allows one to estimate RND from a cross-section of traded options with different strikes. Several alternative techniques for RND estimation have been recently proposed. See Jackwerth (1999) for a literature survey. In this paper, we utilize a new method developed in Bondarenko (2000, 2003b).

In empirical tests, we will group options according to their moneyness. Let $k := K/v_t$ denote the strike-to-underlying ratio, or moneyness. Consequently, a put (call) is

- out-of-the-money (OTM) if $k < 1$ ($k > 1$);
- at-the-money (ATM) if $k = 1$;
- in-the-money (ITM) if $k > 1$ ($k < 1$).

It will be convenient to scale option prices by the value of the underlying. Let $p(k) := P(K)/v_t$ and $c(k) := C(K)/v_t$ denote the *normalized* put and call.

In the absence of arbitrage opportunities, there exists a pricing kernel $m > 0$ such that

$$E[mr_i] = 0, \tag{4}$$

where r_i is the *net* return of a generic security over the holding period $[t,T]$. (That is, the corresponding *gross* return is $R_i = 1 + r_i$.) In particular, $r_p(k)$ and $r_c(k)$ are the net returns on the normalized options, while $r_m = v_T/v_t - 1$ is the net return on S&P 500 (interpreted in this paper as the market portfolio). Recall that all returns are already *in excess* of the risk-free rate and account for S&P 500 dividends.

[3]In reality the risk-free rate is nonzero and S&P 500 does pay dividends. However, in the empirical tests, we convert *spot* prices of all securities into *forward* prices (for delivery at time-T). Forward prices are obtained by adjusting spot prices for the risk-free rate and dividends (when applicable). For example, the forward put price $P(K) = e^{r_f(T-t)} P^s(K)$, where $P^s(K)$ is the spot price and r_f is the risk-free rate over $[t,T]$. When discussing the theory, this convention allows us to abstract from the difference between the S&P 500 value v_t and the S&P 500 futures price F_t. A similar approach has been used in Dumas, Fleming, and Whaley (1998).

2.2 CME Options

Our data consist of daily prices of options on the S&P 500 futures traded on the Chicago Mercantile Exchange (CME) and the S&P 500 futures themselves. The data are obtained from the Futures Industry Institute. The S&P 500 futures have four different maturity months from the March quarterly cycle. The contract size is \$250 times S&P 500 futures price (before November 1997, the contract size was \$500 times S&P 500 futures price). On any trading day, the CME futures options are available for six unique maturity months: four months from the March quarterly cycle and two additional nearby months ("serial" options). The options contract size is one S&P 500 futures. The minimum price movement, or the tick size, is 0.05. The strikes are multiples of 5 for near-term months and multiples of 25 for longer maturities. If at any time the S&P 500 futures contract trades through the highest or lowest strike available, additional strikes are usually introduced.

The sample period in this study is from August 1987 through December 2000. Data for earlier years are not used for two reasons. First, the option market was considerably less liquid during its earlier years. Second, prior to August 1987, options were available only for quarterly maturity months (i.e., only 4 maturities per year). Our analysis requires sets of options which expire each month.

The CME options on the S&P 500 futures and options on the S&P 500 index itself, traded on the Chicago Board Option Exchange (CBOE), have been a focus of numerous empirical studies. For short maturities, prices of the CME and CBOE options are virtually indistinguishable. Nevertheless, there are a number of practical advantages in using the CME options:

- As well known, there is a 15-minute difference between the close of the CBOE markets and the NYSE, AMEX, and NASDAQ markets, where the components of S&P 500 are traded. This difference leads to the non-synchronicity biases between the closing prices of the options and S&P 500. In contrast, the CME options and futures close at the same time (3:15 pm CT).

- It is easier to hedge options using very liquid futures as opposed to trading the 500 individual stocks. On the CME, futures and futures options are traded in pits side by side. This arrangement facilitates hedging, arbitrage, and speculation. It also makes the market more efficient. In fact, even traders of the CBOE options usually hedge their positions with the CME futures.

- Because S&P 500 pays dividends, to estimate RNDs from the CBOE options, one needs to make some assumptions about the Index dividend stream. No such assumptions are needed in the case of the CME futures options.

A disadvantage of the CME options is their American-style feature. However, we conduct our empirical analysis in such a way that the effect of the early exercise is minimal.

Figure 1 provides some descriptive statistics of the data. It plots the average daily trading volume and open interest for different k, when time to maturity $\tau = T - t$ is 1–28 days, 29–56 days, and 57–84 days. The figure illustrates several important features of the data:

(a) The trading activity is relatively light for large τ, but increases considerably as the maturity date approaches. This holds for both trading volume and open interest.

(b) The trading is the heaviest in options with k close to 1. The trading is generally higher in OTM options than in ITM options.

(c) For $k < 1$, puts are more liquid than calls, while the opposite is true for $k > 1$. Among far-from-the-money options, the trading is more active in OTM puts as opposed to OTM calls. This is consistent with the fact that portfolio managers demand OTM puts to hedge their portfolios against stock market declines.

To construct our final dataset we follow several steps, which are explained in Appendix A. In brief, these steps include filtering the option data, forming normalized prices of European puts and calls, and estimating RNDs.

2.3 Overpriced Puts Puzzle

In this subsection we examine historical returns of puts and calls. To build as large a series of non-overlapping returns as possible, we focus on short-term options with one month left to expiration. Table 1 contains a variety of statistics for different strike-to-underlying ratio k. The table is produced in the following way:

- Let j index different options maturities T_j in the sample. We compute returns of options that mature on T_j over the holding period $[t_j, T_j]$, where $t_j = T_{j-1}$. In other words, we consider a rollover trading strategy for which, as soon as one set of options expires, new short-term options are purchased and held until they expire the following month.[4] Overall, there are $N = 161$ one-month holding periods in the sample. (Because there are only 5 option maturities in 1987, N=5+13·12=161.) For each holding period, we compute the net returns $r_p(k)$, $r_c(k)$, and r_m for puts, calls, and the underlying futures.

- On trading date t_j, we classify options according to their strike-to-underlying ratio into equally-spaced bins with centers at $k = 0.94, 0.96, 0.98, 1.00, 1.02, 1.04, 1.06$. Typically, several strikes fall in a given bin. In this case, we select one strike that is the closest to the center of the bin. Thus, there is a maximum of one strike per bin per trading period.

- In Table 1, we report mean, minimum, median, and maximum of $r_p(k)$ and $r_c(k)$ for bins with different k. (In all tables, return statistics are reported as *monthly* and in decimal form. They are *not* annualized.) The pointwise confidence intervals (1%, 5%, 95%, and 99%) are constructed using a bootstrap with 1000 resamples.

  To ensure that the results are not driven by a few extreme returns on low-priced, illiquid options, we do not use option prices lower than 0.5% of the underlying. For example, if v_t=1000, we consider puts and calls no cheaper than \$5 (\$1,250 per option contract). Because of this filter, the number of available returns n may be less than $N = 161$ for some moneyness, especially for OTM options.

The average put return monotonically increases with k. AR is negative and highly significant for all k. In particular, AR is -39% per month for very liquid ATM puts and is -95% per month for less liquid OTM puts with $k = 0.94$. For calls, AR is generally positive but not statistically significant. (The confidence intervals are very wide for OTM calls, reflecting a high variability of their returns.)

Figure 2 provides additional insights by comparing two probability densities: the aggregate risk-neutral density (ARND) and the unconditional objective density (OD). OD is estimated

[4] In practice, options maturity dates are such that $\tau_j = T_j - t_j$ is always either 28 or 35 days. For simplicity, we refer to τ_j as one-month holding period.

using the kernel method from N returns of the underlying.[5] ARND is computed as the point-wise average of N individual RNDs, as functions of moneyness k. Both densities correspond to one-month holding period.

The main differences between the two densities are as follows. Relative to OD, ARND has (i) lower mean, (ii) higher standard deviation, (iii) fatter left tail, and (iv) higher kurtosis. The most pronounced are differences (i) and (iii). In particular, the mean of ARND is lower than the mean of OD by 0.71% (annualized 8.57%). Furthermore, the mean of ARND is located noticeably to the left from its mode. This points to substantial negative skewness of ARND. Differences (i) and (iii) are the primary reasons for negative put returns.

The bottom panel of Figure 2 plots the normalized put and call prices corresponding to ARND and OD. Option prices are obtained by integrating the two densities against option payoffs. The ARND-implied prices may be interpreted as the average option prices over the studied period, while the OD-implied prices may be interpreted as the fair prices computed under the assumptions that 1) investors were risk-neutral, and 2) investors correctly anticipated the distribution of r_m. For all k, the ARND-implied puts are more expensive than the OD-implied ones. In relative terms, the mispricing is the most pronounced for OTM puts, the finding consistent with Table 1. For calls, the ARND-implied prices are lower than the OD-implied prices, except for very high k when the two sets of prices are essentially the same.[6]

To save on space, in what follows we report the empirical results for puts only. For calls, the results are less anomalous and are available upon request.

In Figure 3, we examine whether the results for puts are robust over different subsamples. Specifically, we partition the sample period into four subperiods: 08/87–06/90, 07/90–12/93, 01/94–06/97, and 07/97–12/00. For these subperiods, we report AR for k=0.96, 0.98, 1.00, 1.02, 1.04. For comparison, we also show time-series of the level of the S&P 500 Index and the one-month ATM implied volatility. AR is significantly negative for all subperiods and all k. Predictably, the worst subperiod for selling puts is the first one, which includes the October 1987 market crash. However, even for that "bad" subperiod, AR ranges from -27% to -12% per month for different k. For the next three subperiods, the average returns are generally much lower. As expected, put returns are particularly low in years when the stock market performed well, such as in the second and third subperiods. Typically, AR monotonically increases with k (the only exception being $k = 0.96$ in the first subperiod).

Figure 4 shows the distribution of put returns over time, for three most liquid puts with k=0.98, 1.00, and 1.02. The figure confirms the intuition that put returns exhibit substantial positive skewness. Consider, for example, the ATM put. It expires worthless most of the time, but delivers a high positive return once in a while. The OTM put with k=0.98 has an even more skewed distribution of returns: it expires out-of-the-money even more frequently, but when it does mature in-the-money, returns are more extreme. Figure 4 reveals that skewness increases as puts become more out-of-the-money, which is consistent with the evidence in Table 2.

The fact that puts appear to be overpriced has been noted in a number of recent papers (see the literature review in Section 2.5). In following Sections 2.3.1-2.3.4, we document several new observations which suggest that the *magnitude* of the put mispricing might have been not fully appreciated.

[5]The bandwidth for the kernel method is set to $1.06\hat{\sigma}N^{-1/5}$, where $\hat{\sigma}$ is the sample standard deviation.

[6]For ease of interpretation, the plot shows the no-arbitrage bounds $p(k) \geq (k-1)^+$ and $c(k) \geq (1-k)^+$. These bounds are only relevant for the ARND-implied prices, because the mean of OD is not equal to 1.

2.3.1 Risk-adjustment

First, we examine whether put returns can be justified by some standard asset pricing models. We consider two popular candidates – the Capital Asset Pricing Model (CAPM) and Rubinstein (1976) model. Under CAPM, the pricing kernel depends on the market return r_m as

$$m(r_m) = 1 - \frac{E[r_m]}{Var(r_m)} \left(r_m - E[r_m]\right).$$

Table 2 reports the alpha and beta coefficients for puts with different moneyness k. For any return r_i, alpha and beta are computed as

$$\alpha_i = E[r_i] - \beta_i E[r_m], \qquad \beta_i = \frac{Cov(r_i, r_m)}{Var(r_m)}.$$

CAPM is strongly rejected. For $k \leq 1.0$, the Jensen's alpha is negative and significant at the 1% level. In particular, α is -23% per month for the ATM put and is even lower for the OTM puts. CAPM performs better for high k. This is not surprising, because a long position in a deep ITM put is akin to a short position in the underlying. As expected, put betas are negative and very large in absolute terms, reflecting both negative correlation with the market and substantial leverage. Put betas display a U-shaped pattern with respect to k.

High leverage of puts complicates interpretation of their alpha coefficients. Therefore, Table 2 also reports several risk-adjusted measures that are unaffected by leverage:

- the Sharpe ratio, $SR := \frac{E[r_i]}{\sqrt{Var(r_i)}}$;
- the Treynor's measure, $TM := \frac{\alpha_i}{\beta_i}$;
- M-squared of Modigliani and Modigliani (1997), $M^2 := \frac{E[r_i]}{\sqrt{Var(r_i)}}\sqrt{Var(r_m)}$.

For all k, the Sharpe ratio for selling puts is higher than the Sharpe ratio for the market. The difference is considerable in the case of ATM and OTM puts. The Treynor's measure monotonically increases as k decreases. In economic terms, the Treynor's measure is very large for ATM and OTM puts. Similarly, ATM and OTM puts appear substantially overpriced according to the M^2 measure. (Intuitively, M^2 shows the return that an investor would have earned if a particular position had been diluted or leveraged to match the standard deviation on the market portfolio.)

It is well-known that the Sharpe ratio and related measures can be misleading when returns exhibit substantial skewness. See, for example, Goetzmann, Ingersoll, Spiegel, and Welch (2002). Therefore, as alternative risk-adjustment, we now consider the Rubinstein model. In this model, the pricing kernel depends on the market return r_m as

$$m(r_m) = Const \cdot \frac{1}{(1 + r_m)^{\gamma}}, \tag{5}$$

where $\gamma > 0$ is the coefficient of relative risk aversion of the representative investor. For different k, we find the coefficient γ such that

$$E\left[r_p(k)\,\frac{1}{(1 + r_m)^{\gamma}}\right] = 0.$$

The results are reported in Table 2. The important observation is that no single γ can simultaneously explain put returns across all levels of moneyness. Generally, γ increases as k decreases; γ ranges from a reasonable value of 4.3 for deep ITM puts to a very large value of 131 for deep OTM puts. For comparison, γ for the market return is 4.3.

The specification in (5) has also been investigated in Coval and Shumway (2001). They study zero-beta straddles on the S&P 500 index options (CBOE) with moneyness k close to 1.0. Using weekly returns and the sample period from 01/1990 to 10/1995, Coval and Shumway report the estimates of γ from 5.68 to -6.68.

2.3.2 Extreme returns

In Table 3 we study how sensitive the overpriced puts puzzle to extreme observations by examining periods with the highest put returns in the sample. These periods correspond to option maturities in the following five months:

(a) 10/87 (precedes the October 87 crash),

(b) 11/87 (includes the October 87 crash),

(c) 08/90 (includes the August 90 crash – the invasion of Kuwait by Iraq),

(d) 04/94,

(e) 08/98 (includes the August 98 crash – the Russian debt default).

It is somewhat surprising that period (b), which includes the October 87 crash, was not the worst month for selling puts – in fact, it was only the forth worst after periods (a), (c), and (e). Even though the decline in the underlying was the largest (-14%) over period (b), puts were selling at unusually high prices at the beginning of the period (as evidenced by the corresponding ATM volatility in Table 3). The market was very volatile and put premiums were high because the S&P 500 had already fell substantially in the previous month. In periods (a), (c), and (e), the returns in the underlying were less dramatic (-11%, -10%, -9%). However, they happened after relatively calm periods, when puts were inexpensive by historical measures.

It is clear from Table 3 that put sellers may occasionally incur huge losses. One could argue, therefore, that if these extreme losses had happened in the sample more frequently, then the profitability of selling puts might have disappeared.

To explore this possibility, we compute how many extreme observations must be added to the empirical distribution of put returns to make the average return become zero. Specifically, for each k we add to the sample $l = l(k)$ identical returns corresponding to the October 87 crash (i.e., period (b) above) so that the new average return is zero. The results are reported in Table 4. For example, about l=18 additional October 87 returns (=346%) are needed for ATM puts to break even. This corresponds to about 1.3 crashes a year![7]

We repeat the same exercise but now add artificial returns equal to the *highest* return from periods (a)-(e), which may be different for different k. For example, the highest ever return

[7]Jackwerth (2000) conducts a similar analysis. He studies returns for the S&P 500 index option (CBOE) over the period from 10/1988 to 12/1995 and finds that, in order to make the alpha coefficient for selling the ATM put and OTM put (k=0.95) equal zero, artificial 20% crashes have to be added one in about every 4 years. The results in Table 4 seem to be even more extreme. To make AR equal zero, crashes of the October 87 magnitude have to happen one in about 9 months.

of the ATM put was in period (c) (=540%) and 12 such returns must be added to the sample before the average return becomes zero.[8]

2.3.3 Bull market?

Buying put options is a *bearish* strategy. That is, put returns are low when the market performs well and vice versa. Over the sample period, the level of the S&P 500 Index has risen more than 4 times, from v_0=314.59 to v_{T^*}=1312.15. Is it possible, therefore, that selling puts was so profitable simply because of the unprecedented bull market of the late nineties? In other words, maybe selling puts would not work in downward trending markets?

To explore this possibility, we perform another exercise. We introduce a negative drift for the S&P 500 Index and compute the value of the drift that would reconcile the historical put returns. Specifically, we replace the true process for the S&P 500 Index v_t with the modified process $\hat{v}_t = v_t e^{-\eta t}$ for some constant $\eta > 0$. This implies that the return on the underlying futures r_m over the period $[t, T]$ is reduced to $1 + \hat{r}_m = (1 + r_m)e^{-\eta(T-t)}$. Using the modified returns on the underlying, we then recompute put returns for all holding periods as well as the average returns. For each k, the average return monotonically and continuously increases as η increases. This observation allows us to find the critical value of η that makes the average return equal zero. The results are reported in Table 5, which reveals that a negative drift of -1.5% per month is necessary for the ATM put to break even. Assuming this drift, the final value of the S&P 500 Index at the end of the sample period v_{T^*} would have been only 111.9 instead of 1312.15! For 2% and 4% OTM puts, the necessary drifts are -2.0% and -2.7% per month, with the corresponding final values of S&P 500 being only 47.8 and 13.3, respectively.

(Intuitively, Sections 2.3.2 and 2.3.3 look at different characteristics of the empirical distribution for r_m and $r_p(k)$. In the former, we modify the empirical distribution of put returns by increasing the frequency of most influential observations. In the latter, we shift the mean of the distribution for the market return to the left, without changing the distribution's higher moments. This is consistent with the intuition in Merton (1980), who points out that estimating the mean of the empirical distribution is more difficult than estimating the standard deviation. The latter approach gradually increases *all* put returns, until the condition AR=0 is satisfied.)

Overall, Table 5 implies that one would need to introduce a highly implausible drift to justify historical put returns. This also suggests that the exceptional bull market of the nineties cannot be the main reason for the put puzzle. In fact, the mispricing that can be attributed to the bull market is likely to be very small. For example, even if we choose the drift η so that the risk premium of the S&P 500 futures over the 14-year period is zero (i.e., the market earns on average just the risk-free rate), then monthly AR for puts with $k = 0.94, \ldots, 1.06$ are still very low: -0.89, -0.49, -0.41, -0.21, -0.09, -0.03, and -0.003, respectively.

2.3.4 Wealth transfer

The economic impact of the put mispricing is likely to be considerable, due to high trading volume of the S&P 500 options. We can obtain a rough estimate of the economic impact by

[8] It is important to stress that the exercises in this and the following subsections are not meant to imply that fair put returns *should* be zero on average. After all, selling puts *is* risky and must be rewarded with risk-premium. Instead, the purpose of the exercises is to assess intuitively whether a particular explanation has the potential to generate a high mispricing of puts.

computing the total profit or loss (P&L) from holding long put positions. For $t < s$, let $PL_{t,s}$ denote P&L from all put positions over the period $[t, s]$:

$$PL_{t,s} = PL_{t,t+1} + PL_{t+1,t+2} + \ldots + PL_{s-1,s},$$

where $PL_{t,t+1}$ denotes the daily profit or loss. To compute $PL_{t,t+1}$, we aggregate P&L of individual puts with all available strikes and all maturities from 1 to 365 days. (P&L of a particular put is the price change over $[t, t+1]$ times the time-t open interest, or $(P_{t+1} - P_t)O_t$.)

Intuitively, $PL_{t,s}$ shows the amount of wealth transfered between put buyers and put sellers over the period $[s, t]$, subject to two simplifying assumptions that 1) all traders can be divided into either buyers (who only hold long puts) or sellers (who only hold short puts),[9] and 2) each day, options are traded at the settlement prices. We report the cumulative wealth transfer $PL_{t,s}$ for the entire period, as well as the four subperiods:

	08/87–06/90	07/90–12/93	01/94–06/97	07/97–12/00	Full sample
PL:	-0.1 bln	-2.3 bln	-6.5 bln	-8.9 bln	-17.8 bln

Over the whole period, put buyers have lost to put sellers astounding \$17.8 bln.[10] However, the economy-wide impact of the put mispricing is likely to be even larger considering that

- in addition to the CME options, there exist a number of other options based on broad market indexes. CBOE, CME, CBOT, and other exchanges list options on various indexes, their futures, and related Exchange-Traded Funds (ETFs). Trading in many of these contracts is very active, including options based on S&P 100, S&P 500, S&P MidCap 400, Russel 2000, DJIA, NASDAQ 100, and others.
- in addition to options on market indexes, there are numerous options on individual stocks. Puts on individual stocks also appear overpriced, although to a lesser extent.
- besides organized exchanges, considerable amount of equity options is traded over-the-counter. OTC transactions often involve contracts with longer maturities.

2.4 Robustness of Findings

The findings reported in Section 2.3 are not sensitive to a variety of checks in the empirical methodology. In particular, the results are not affected when we 1) use option closing prices instead of settlement prices, and 2) modify the filtering criteria. In an earlier version, we excluded year 1987 from the sample. Naturally, the exclusion of the October 87 crash has the effect of making the average put returns even lower, however, not *much* lower. As follows from Section 2.3.2, the put anomaly is not driven by a few extreme observations.

It should be also noted that the results cannot be explained by the transaction costs or bid-ask spreads. This is because we focus on buy-and-hold strategies that involve very little trading. In fact, options are assumed to be traded only once, at the beginning of each period. At the end of each period, options either expire worthless (which happens most of the time)

[9] In reality, some traders hold both long and short positions when, for example, creating put spreads. This assumption has the effect of exaggerating the estimates of wealth transfer from put buyers to put sellers.

[10] Note that traders can also create *synthetic* put positions via put-call parity. For example, an ATM call can be used to create a position equivalent to an ATM put.

or are exercised at known prices (with a small commissions for the exercise). Because the magnitude of the mispricing of puts is so large, introducing reasonable market imperfections (trading costs, bid-ask spreads, price impact, costs associated with maintaining the margin requirements, etc.) have a relatively small effect on the average returns. The only exception might be extremely deep OTM puts. Recall, however, that we do not use very low-priced, illiquid options.

The findings reported in Section 2.3 are not specific to a particular choice of the holding period. When we repeat the previous analysis with time to maturity τ=3 months, the general findings are qualitatively similar to the case of τ=1 month. The average put returns are negative for all moneyness. For ATM and OTM puts, the results are very significant, both economically and statistically.

2.5 Related Literature

Several recent papers have documented related findings, including Jackwerth (2000), Coval and Shumway (2001), Aït-Sahalia, Wang, and Yared (2001), Bakshi and Kapadia (2003), and Bollen and Whaley (2003). These papers look at different trading strategies and datasets, but the general conclusion is that puts (especially ATM and OTM) have been historically too expensive. Noteworthy, some papers use transactions data and find that the transaction costs and bid-ask spreads have little effect on monthly put returns (see Coval and Shumway (2001), Bollen and Whaley (2003)).

It is common in the literature to study the profitability of the so-called "crash-neutral" strategies (Jackwerth (2000), Coval and Shumway (2001)), where a deep OTM put is used to limit losses in the case of market crashes. For example, consider a position which is short the ATM put and long the OTM put with $k = 0.90$, that is, $Z_t = p_t(0.90) - p_t(1.00)$. The OTM put limits the downside risk, with the lowest terminal value Z_T being -10% of the index's initial value v_t. The position earns the same return whether the market return is -10% or -50%.

We want to point out that this approach implicitly assumes a very specific way to risk-adjust future payoffs. To see this more clearly, consider strategies that have capped payoff for market declines below some critical value v^c. Specifically, let Z_t^c denote the value of a "crash-neutral" strategy for which payoff $Z_T^c = 0$ when $v_T \leq v^c$. (The normalization of the payoff to zero in crash states is without loss of generality when the risk-free bond is available.) The average excess return on all such strategies will be zero if and only if the pricing kernel has the form:

$$m_T = m(v_T) = Const, \qquad \text{all } v_T \geq v^c.$$

In other words, investors are effectively assumed to be risk-neutral over the range of values $v_T \geq v^c$. In particular, if v^c is set to $0.9v_t$, then investors are indifferent between payoffs received when the market return is -5% or +15%. In Section 3, we will test for these and much more general pricing kernels.

2.6 Alternative Explanations

In the rest of the paper, we will implement a new methodology to explore three explanations of the overpriced put puzzle.

E1: Risk premium

According to this explanation, high prices of puts are *expected* and reflect normal risk premiums under some equilibrium model. From Section 2.3.2, we know that the canonical models, such as CAPM and Rubinstein (1976) model, cannot explain the empirical findings. Still, maybe there is another, nonstandard model which can. In this *true* model, investors must strongly dislike crash-like returns of S&P 500 and are willing to pay considerable premiums to for puts that provide explicit insurance against market declines.

E2: The Peso problem

According to this explanation, the sample under investigation is affected by the Peso problem.[11] That is, in spite of including the October 1987 crash, investors *correctly* anticipated more market crashes over the 14-year period but those did not happen. In this case, the *ex post* realized returns of S&P 500 are different from investors' *ex ante* beliefs. Puts only *appear* overpriced, and the mispricing would have disappeared if data for a much longer period were available. Section 2.3.2 suggests that unrealized crashes are unlikely to explain *all* mispricing of puts. Still, it is possible that the Peso problem is responsible for a *portion* of the anomaly.

It should be mentioned that the Peso problem is often defined narrowly to arise when the distribution of the data generating process includes a catastrophic state that occurs with a very low probability. Because this state has low probability, it may not be observed in a given small sample. Because the state is catastrophic, the possibility of this state occuring substantially affects equilibrium prices. Here, we understand the Peso problem more broadly as arising whenever the *ex post* frequencies of states within the data sample differ considerably from their *ex ante* probabilities, and where these deviations distort econometric inference. In other words, the Peso problem is present when the sample moments calculated from the available data do not match the population moments that investors use to make their decisions.

E3: Biased beliefs

According to this explanation, investors' subjective beliefs are mistaken. Similar to E2, this explanation states that the S&P 500 realized returns have not been anticipated by investors. The OTM puts were expensive because investors assigned too high probabilities to negative returns of S&P 500. Perhaps, memories of the 1987 stock market crash were still fresh and, even though the *true* probability of another extreme decline was small, investors continued to overstate this probability.

3 Model-Independent Approach

In this section, we implement the model-free methodology for testing rationality of asset pricing. We start by reviewing the new theory developed in Bondarenko (2003a).

3.1 New Restriction on Securities Prices

Suppose that securities are traded in a frictionless and competitive market. As before, let Z_t denote the value of a generic security with a single payoff Z_T at time-T. The payoff Z_T may be path-dependent. The risk-free rate is normalized to zero.

[11]The Peso problem is analyzed in, for example, Bekaert, Hodrick, and Marshall (1995).

Let $E_t[\cdot]$, $E_t^S[\cdot]$, and $E_t^N[\cdot]$ denote the time-t conditional expectations under the *objective*, *subjective*, and *risk-neutral* probability measures. The objective measure reflects the *true* (or *large-sample*) frequencies with which various events occur. The subjective measure represents investors' *beliefs* regarding future events. The risk-neutral measure always exists if the market is arbitrage-free (Harrison and Kreps (1979)) and is unique if the market is complete (Harrison and Pliska (1981)). However, we do not insist on market completeness. Securities prices can be computed under the risk-neutral measure as $Z_t = E_t^N[Z_T]$. Alternatively, prices can be expressed under the subjective measure using the pricing kernel as

$$m_t Z_t = E_t^S[m_T Z_T], \qquad m_t = E_t^S[m_T].$$

The Efficient Market Hypothesis (EMH) is characterized by two conditions: (i) rational learning, which means that, when new information arrives to the market, investors update their beliefs using the rules of conditional probability, and (ii) correct beliefs, which means that the objective and subjective measures coincide, or $E_t[\cdot] = E_t^S[\cdot]$. Following Bossaerts (2003), we allow for a more general case of the *Efficiently Learning Market* (ELM). Bossaerts argues that of the two conditions underlying EMH, it is the condition of rational learning that reflects the essence of rationality. In ELM, he maintains (i), but relaxes (ii).

Under EMH, the security's price must satisfy the following standard restriction:

$$E_t[Z_s m_s] = Z_t m_t, \qquad t < s. \tag{6}$$

To test the restriction in (6), one must pre-commit to a specific pricing kernel. As a result, empirical tests suffer from a joint hypothesis problem: rejections may emerge because the market is truly inefficient or because an incorrect pricing kernel has been assumed.

Bondarenko (2003a) shows that, under certain conditions, there is another martingale-type restriction on securities prices. This restriction is stated in Proposition 1 below. In order to give our empirical findings the broadest interpretation possible, we prove a more general version of the result in Bondarenko (2003a). Let x_T denote some general state variable. Denote $E_t^x[\cdot] = E_t[\cdot\,|\tilde{x}_T = x]$ and $E_t^{S,x}[\cdot] = E_t^S[\cdot\,|\tilde{x}_T = x]$ as the objective and subjective expectation *conditional* on the final state being x. Also, denote $f_t(x_T)$, $g_t(x_T)$, and $h_t(x_T)$ as the objective, subjective, and risk-neutral densities of x_T. Finally, let $m_t^*(x_T) := E_t^{S,x_T}[m_T]$ denote the *projection* of the pricing kernel onto the final state. (For discussion of projected kernels, see, for example, Hansen and Richard (1987), Cochrane (2001), and Rosenberg and Engle (2002).) The projected kernel $m_t^*(x_T)$ has the property that at time-t it correctly prices securities whose time-T payoffs depend on x_T, or $Z_T = Z(x_T)$.

In what follows, we fix three dates $t < s < T$ and study securities returns over the period $[t, s]$. We say that the projected kernel is *path-independent* if the following assumption holds.

Assumption 1 *For all histories and all x_T, $m_t^*(x_T) = m_s^*(x_T)$.*

Bondarenko (2003a) focuses on a special case of Assumption 1 where the pricing kernel m_T is an arbitrary function of the state variable x_T:

$$m_T = m(x_T). \tag{7}$$

Under specification (7), Assumption 1 is satisfied trivially, because $m_t^*(x_T) = m(x_T)$ for all t. However, the assumption is more general than (7). In particular, m_T could depend on other state variables besides x_T, or x_T might be an imperfect proxy for the true state variable. We will discuss Assumption 1 in more detail in Section 4.

Proposition 1 *Suppose that ELM and Assumptions 1-3 hold. Then securities prices satisfy*

$$E_t^x\left[\frac{Z_s}{h_s(x)}\right]=\frac{Z_t}{h_t(x)}. \tag{8}$$

Two additional Assumptions 2-3 and the proof of Proposition 1 are given in Appendix B. Proposition 1 extends the main result in Bondarenko (2003a) by replacing condition (7) with more general Assumption 1. Intuitively, the restriction in (8) says the following. Suppose that an empiricist observes many repetitions of the same environment and selects only those price histories for which $\tilde{x}_T = x$. Then, in the selected histories the ratio $Z_t/h_t(x)$ must change over time unpredictably. The unusual feature of (8) is conditioning on future information. The expression inside the expectation operator in (8) is not know at time-s and it could only be computed after the final state at time-T is revealed.

The scope of the new restriction is quite general. Bondarenko (2003a) argues that essentially every known equilibrium model has the property that its pricing kernel satisfies (7) for some simple state variable x_T, with kernel m_T depending on the state x_T but not the complete history $(x_t : t \leq T)$ for each time-T. For example, this holds for

- CAPM and Rubinstein (1976) model for which $x_T = v_T$, where v_T is the value of the market portfolio;
- the canonical consumption-based models for which $x_T = c_T$, where c_T is aggregate consumption (or, a vector of several consumption goods);
- Epstein and Zin (1989, 1991) and Weil (1989) models with recursive preferences for which $x_T = (v_T, c_T)$;
- the habit formation models of Abel (1990) and Campbell and Cochrane (1999) for which $x_T = (c_T, \omega_T)$, where ω_T is external habit;
- the multifactor arbitrage models for which x_T is a vector of common factors.

Specification in (7) has the important implication that certain trading strategies – termed *statistical arbitrage opportunities* (SAOs) and defined with respect to the state variable x_T – cannot exist in equilibrium. The absence of SAOs then implies the restriction in (8). Note that, for different choices of the state variable x_T the restriction in (8) allows one to test different classes of equilibrium models.

3.2 Discussion

The new restriction in (8) has three important properties. First and most significantly, the restriction makes no reference to the unobservable pricing kernel. Instead, the restriction requires the knowledge of RND $h_t(x)$, which is implicit in securities prices and can be estimated if options on x_T are traded. This means that the restriction in (8) can be used to resolve the joint hypothesis problem present in tests of EMH. It allows one to test whether securities prices are compatible with *any* equilibrium model, for which the pricing kernel satisfies Assumption 1.

Second, the restriction in (8) can be used in small samples and in the presence of selection biases with respect to x_T. To see this more clearly, suppose that the empiricist has collected a dataset of price histories with final states x^j, $j = 1, \ldots, N$. As the number of histories N increases to infinity, the corresponding empirical density $\hat{f}_t(x_T)$ will approach the objective density $f_t(x_T)$. For small N, however, the empirical density $\hat{f}_t(x_T)$ might differ from $f_t(x_T)$ considerably. This will usually cause rejection of the restriction in (6), even if the true pricing kernel were known. However, because the restriction in (8) applies for *every* realization x_T, the restriction will hold even in small samples. As another illustration, suppose that the empiricist has to use a dataset in which certain realizations x_T are explicitly excluded. Specifically, suppose that the dataset only contains those histories for which $x_T \in A$, where A is a subset of possible final states. Again, under such a selection bias, the restriction in (6) will normally be rejected. Interestingly, the selection bias does *not* affect the restriction in (8) – because of conditioning on the final outcome, the restriction holds for *any* subset A.

Third, the new restriction continues to hold even when investors have incorrect beliefs about distribution of x_T. Specifically, suppose that investors start with expectations $g_t(x_T) \neq f_t(x_T)$, but they update their expectations using Bayes' law and correct likelihood functions (see Assumption 2 in Appendix B). Then, the restriction in (8) must still hold.

The new restriction can be implemented in either *parametric* or *nonparametric* contexts. In the former, one builds RND $h_t(x)$ from a parametric equilibrium model. In this approach, there is no restriction on how broad and general the state variable x_T could be. Although the approach will suffer from the same joint hypothesis problem that affects empirical tests based on the standard restriction in (6), the approach still could be useful. First, the new restriction offers an alternative way to test asset pricing (which has been largely overlooked in the literature). Thus, there might exist situations when the standard restriction is not rejected while the new one is. Second, the new restriction still possesses the other two important advantages (i.e., robustness to selection biases and distorted beliefs).

Still, it is the nonparametric context where the new restriction is probably the most useful. In this case, one makes no strong assumptions about the true equilibrium model/pricing kernel and estimates RND $h_t(x)$ from traded securities. This approach, therefore, is best suited to situations where the state variable x_T corresponds to prices traded assets for which liquid option markets exist. In Section 3.3, we implement this approach for $x_T = v_T$, the value of the market portfolio. The restriction in (8) now becomes

$$E_t^v \left[\frac{Z_s}{h_s(v)} \right] = \frac{Z_t}{h_t(v)}, \tag{9}$$

The restriction in (9) holds for all equilibrium models for which the projection of the pricing kernel on v_T is path-independent, or

$$m_t^*(v_T) = m_s^*(v_T), \tag{10}$$

where $m_t^*(v_T) := E_t^{S,v_T}[m_T]$ and $m_s^*(v_T) := E_s^{S,v_T}[m_T]$. A special case of this specification is when the pricing kernel is an arbitrary function of v_T:

$$m_T = m(v_T). \tag{11}$$

One simple setting for which the restriction in (9) applies is the following. Consider a pure-exchange economy with a finite horizon. There are one risky asset (the market) and a risk-free bond. A representative agent maximizes the expected value of utility function $E_t[U(v_T)]$, with $U' > 0$ and $U'' \leq 0$. Since the pricing kernel in this economy is $m_T = U'(v_T)$, condition (11) is satisfied. Note that this holds true no matter how complex the process for v_t is (which, in particular, could include stochastic volatility, jumps, and multiple factors).

To gain some intuition for the model-free methodology, Appendix C presents a parametric example, which illustrates the properties of the new restriction in the presence of 1) risk-aversion, 2) incorrect beliefs, and 3) selection biases.

3.3 Test of New Restriction

To test the restriction in (9), we first rewrite it in a slightly different form. Suppose that three dates $t < s < T$ are fixed and let $\lambda^v := h_t(v)/h_s(v)$ denote the inverse of the return of RND evaluated at the final value v. Applying (9) to the risk-free bond $Z_t \equiv 1$, we obtain that

$$E_t^v[\lambda^v] = 1. \tag{12}$$

In view of (12), the restriction in (9) can be expressed as

$$E_t^v[\lambda^v r_i] = 0, \tag{13}$$

where $r_i = Z_s/Z_t - 1$ is the net return over $[t, s]$. In this form, the restriction in (13) resembles the restriction in (4). Intuitively, the strictly positive random variable λ^v plays the role of a relevant "discount factor" for the expectation $E_t^v[\cdot]$ as does the pricing kernel m for the expectation $E_t[\cdot]$. By taking unconditional (over time) expectations of (13), we obtain

$$E\left[E_t^v[\lambda^v r_i]\right] = 0. \tag{14}$$

In moment (14), time is integrated out, but conditioning on the future value is still present. Intuitively, for *every* possible realization v, the random variable $\lambda^v r_i$ must have zero mean.

To test the condition in (14), we proceed as follows. As in Section 2.3, let j index option maturities T_j. We again compute monthly returns r^j over the holding period $[t_j, s_j] = [T_{j-1}, T_j]$. That is, we consider non-overlapping monthly returns from one option maturity date to another. For conditioning on the future information, we use the next option maturity date T_{j+1}. Therefore, $\lambda^j = h_{t_j}(v_{T_{j+1}})/h_{s_j}(v_{T_{j+1}})$. For each holding period, we estimate two RNDs from options that mature at date T_{j+1}. (These options have 2 months left to maturity at the beginning of the period t_j and 1 month left to maturity at the end of the period s_j.) We use only those holding periods $[t_j, s_j]$ for which 1) necessary option series are available to estimate both RNDs, and 2) the final value $v_{T_{j+1}}$ falls within the lowest and the highest strikes available on both trading dates t_j and s_j, which ensures that accurate estimation of both $h_{t_j}(v_{T_{j+1}})$ and $h_{s_j}(v_{T_{j+1}})$ is possible. Overall, there are now 144 usable holding periods $[t_j, s_j]$.

Armed with weights λ^j, we form the average *weighted* return (AWR)

$$\text{AWR} = \frac{1}{n}\sum_j^n \lambda^j r^j,$$

where n is the number of available weighted returns. AWR is the sample analogue of $E[E_t^v[\lambda^v r_i]]$. The condition in (14) says that AWR should be insignificantly different from zero.

Although the expression for AWR might appear rather "conventional," it is important to reiterate that its weights are constructed by using *future* information. Specifically, the weight λ^j depends on the value $v_{T_{j+1}}$, not yet known at time-T_j. (In probability theoretical terms, the weight λ^j is a random variable which is *not* measurable with respect to information set at time-T_j.) This means that computing AWR requires a particular look-ahead bias. This look-ahead bias would normally present a serious problem for traditional empirical tests, but the bias is the very reason why our model-free approach works.

As in Section 2.3, we focus on one-month puts with different moneyness. Specifically, we compute the return $r_p^j(k)$ over the holding period $[t_j, s_j]$ on put with maturity s_j, where k is moneyness on trading date t_j. Table 6 reports mean, minimum, median, and maximum of weighted return $\lambda r_p(k)$, for different k. The pointwise confidence intervals (1%, 5%, 95%, and 99%) are constructed using a bootstrap. The left panel of Figure 5 summarizes the main results of Table 6 by plotting AWR as a function of k. Also shown are the 5% and 95% confidence intervals. We find that AWR is negative for all k and that it is statistically significant at the 5% significance level for all $k \leq 1.02$ and at the 1% level for all $k \leq 1.00$.

Since the condition in (14) must hold for many other securities and trading strategies Z_t, we can test it for two additional cases:

(i) r_i is the return on S&P 500 futures, that is, $Z_t \equiv v_t$;

(ii) r_i is the return on a *two*-month put as opposed to a *one*-month put.

In both cases, we compute returns over the same monthly holding periods $[t_j, s_j] = [T_{j-1}, T_j]$ as before and use previously computed weights λ^j to form the weighted return. In case (i), we find that AWR for S&P 500 is positive (AWR=0.50%). However, it is not statistically significant (t-statistics is 1.09).

Case (ii) corresponds to the rollover trading strategy that buys puts 2 months before maturity and sells them 1 month before maturity. Puts are again classified according to their moneyness on date t_j with k=0.92, 0.94, ..., 1.08. The main findings can be summarized as follows. AR is negative for all k: -44%, -39%, -38%, -33%, -27%, -21%, -16%, -12%, and -11%, respectively. AR is statistically significant at the 5% significance level for all $k \leq 1.02$ and at the 1% level for all $k \leq 1.00$. As expected, average returns for two-month puts are less extreme than those for one-month puts. (See Table 1.) As for AWR, it is also negative for all k: -38%, -36%, -33%, -28%, -22%, -16%, -12%, -8%, and -6%, respectively. AWR is statistically significant at the 5% significance level for all $k \leq 1.02$ and at the 1% level for all $k \leq 1.00$. AWR for two-month puts and the confidence intervals are shown in the right panel of Figure 5.

Overall, the results in this section imply that *no* equilibrium model with a pricing kernel satisfying (10) can possibly explain the put anomaly, even when allowing for the possibility of the Peso problem and incorrect beliefs.[12]

4 Interpretation of Results

This section discusses in more detail what possibilities are ruled out by the empirical results of Section 3.3.

4.1 Rejected Models

The results in Section 3.3 rule out the whole class of equilibrium models. In those models, pricing kernels are restricted to the form (11). At the same time, rejected models can have very general price dynamics for v_t, including those with jumps and/or stochastic volatility. In

[12] All results in this section are robust to the variations in the empirical design discussed in Section 2.4.

other words, the empirical results make a statement about the pricing kernel, not the data generating process.

To clarify this point, let $\mathcal{D}$ denote the class of price processes which are supported by a pricing kernel satisfying (11), for a given horizon T. Consider a continuous-time pure-exchange economy as in Appendix C. There are a risky asset (the market) v_t and a risk-free bond. The risk-free rate is to zero. Also traded are various derivative securities in zero net supply. The representative investor maximizes the expected value of $U(v_T)$. For simplicity, we assume the CRRA preferences in (17), so that $m_T = v_T^{-\gamma}$.

Suppose that investors receive information about the terminal value v_T represented by an exogenous process ψ_t, with $\psi_T = v_T$. For any process ψ_t, the price dynamics v_t is endogenously derived by solving the representative investor's portfolio problem. Different choices for the process ψ_t imply different processes v_t. For example, suppose that ψ_t follows a Geometric Brownian motion:

$$\frac{d\psi_t}{\psi_t} = \mu dt + \sigma dB_t,$$

where μ and σ are constant. Then, v_t also follows a Geometric Brownian motion:

$$\frac{dv_t}{v_t} = \gamma\sigma^2 dt + \sigma dB_t,$$

with $v_t = \psi_t \exp((\mu - \gamma\sigma^2)(T - t))$. This is, of course, the case of the Black-Scholes model.

Suppose next that ψ_t follows a jump-diffusion:

$$\frac{d\psi_t}{\psi_{t-}} = (\mu - \lambda\mu_J)dt + \sigma dB_t + dN_t,$$

where N_t is a Poisson jump process with arrival intensity λ and stochastic jump size e^{y_i}. Processes B_t and N_t are independent. For each jump i, y_i is normally distributed with mean $(\mu_y - 0.5\sigma_y^2)$ and variance σ_y^2. The expected jump size is $\mu_J = e^{\mu_y} - 1$. It follows from Naik and Lee (1990) that the equilibrium price process v_t now is also a jump-diffusion:

$$\frac{dv_t}{v_{t-}} = (\gamma\sigma^2 - \lambda c)dt + \sigma dB_t + dN_t,$$

and $v_t = \psi_t \exp\left((\mu - \lambda\mu_J - \gamma\sigma^2 + \lambda c)(T - t)\right)$.[13]

The information arrival process ψ_t can be made even more general. There could be multiple fundamental factors (including multi-factors stochastic volatility and general jumps in price and volatility). There could be arbitrary correlations between fundamental factors, and parameters could be time-varying. Obviously, such a general specification will not admit an analytical solution for the endogenously determined price process v_t. However, the solution (whenever exists) can still be found via numerical methods. The price process will inherit general properties of the process for ψ_t. Similar to ψ_t, the price v_t will follow a multi-factor process with jumps and stochastic volatility. In this setting, the objective price process, a derivative's price Z_t, the risk-neutral density $h_t(v_T)$ will all depend not only on v_t but other factors. Nevertheless, because the pricing kernel is $m_T = v_T^{-\gamma}$, even these very flexible specifications are in the class of rejected models $\mathcal{D}$.[14]

[13] The constant c is given by $c = \exp\left((1-\gamma)\mu_y - 0.5\gamma(1-\gamma)\sigma_y^2\right) - \exp\left(-\gamma\mu_y + 0.5\gamma(1+\gamma)\sigma_y^2\right)$.

[14] Admittedly, the described general specifications are quite complex and intractable. However, there is no reason to believe that the *true* model is simple and tractable.

Although we have focused on one utility specification, the standard CRRA preferences, the whole point is that $U(v_T)$ could be an arbitrary function, corresponding to various preferences. Moreover, additional price dynamics in $\mathcal{D}$ obtain by considering alternative general equilibrium constructions (not just pure-exchange economies with terminal consumption).[15]

Our approach also rules out equilibrium models for which the pricing kernel m_T depends on other relevant state variables besides v_T, provided that the projected kernel $m_t^*(v_T)$ is path-independent as stated in (10). This has two important consequences. First, since our empirical application focuses on very short horizons, many state variables identified in the theoretical literature are not likely to be important. In particular, the aggregate consumption c_t and habit ω_t discussed in Section 3.1 have very smooth time series at the monthly frequency, especially when compared to v_t. As a result, these state variables are not able to introduce meaningful path-dependence of the projected kernel $m_t^*(v_T)$. Second, many theoretical models use the market portfolio as the relevant state variable. In this paper, we approximate the market portfolio with the S&P 500 index. Although standard in the empirical literature, this approach is open to the Roll's critique. However, because Proposition 1 allows us to focus on the projected kernel $m_t^*(v_T)$ instead the original kernel, we do not have to worry that the S&P 500 value v_T might be an imperfect proxy of the true market portfolio.

To be able to explain the put puzzle, a candidate equilibrium model must produce a projected kernel $m_t^*(v_T)$ which is *strongly* path-dependent, considering how little can be explained when path-independence is assumed. However, such models have not received much attention in the literature. In particular, currently there is no accepted general equilibrium model where the representative investor's utility function $U(\cdot)$ explicitly depends on stochastic volatility.

4.2 Peso Problem and Mistaken Beliefs

In addition to rejecting the broad and important class of price dynamics, the results in Section 3.3 also rule out the explanations E2 and E3. No previous paper has studied the *combined* effect of risk-aversion, selection bias, and mistaken beliefs. To see why this is significant, consider the parametric example in Appendix C. The example assumes the CRRA preferences with $\gamma = 4$ and a single-factor diffusion for v_t. It demonstrates that, for an arguably plausible combination of risk-aversion, mistaken beliefs, and the selection bias, it could be possible to produce a realistic mispricing of puts, where the mispricing is about of the same magnitude and shape as that observed empirically (see Figure 6). At this point, one might even incorrectly conclude that the put puzzle is solved, for a simple and parsimonious model can achieve almost a perfect fit to the data. However, the candidate "solution" does not survive a test of the new restriction. When the new restriction is applied, the mispricing in the parametric example "disappears", while the mispricing in the data does not. This means that the mispricing in the data is of *different* origin and is not due to risk-aversion, beliefs, and the selection bias.

In this respect, our paper offers two important insights. First, the biggest problem with put options is not the *magnitude* of the mispricing – a very substantial mispricing could be generated in plausible settings with reasonable parameters. Instead, the *real* puzzle is *intertemporal inconsistency* of put prices, as evidenced by the rejection of the new restriction. The latter is equivalent to saying that there exist statistical arbitrage opportunities. Second, to falsify spurious "solutions" of the put puzzle, one might need to test the new restriction. Even

[15] It is beyond the scope of this paper to formally characterize price processes in $\mathcal{D}$. For the special case of pure-exchange economies and when v_t follows a single-factor Markov diffusion, the relevant results are available in Bick (1990) and He and Leland (1993).

though a lot of flexibility can be achieved by varying preferences, beliefs, and selection bias, still, securities prices cannot be completely arbitrary and have to respect the new restriction.

4.3 Not Rejected Partial-Equilibrium Models

Although the class of rejected price dynamics is quite rich, we want to make it clear that there are also many models that are *not* in this class. Important examples are models of Heston (1993), Hull and White (1987), Bates (2000), and related models. It is important to point out that all these models are *partial equilibrium* (PE) models. As a primitive assumption, they assume a parametric price dynamics for v_t as well as market prices of various risk factors. These models do not address the issue whether the assumed price dynamics could be supported by some economically sensible preferences of the representative investor. Because price dynamics could be essentially arbitrary, pricing kernels in PE specifications do *not* usually satisfy (11).

As *reduced-form* approximations, PE models can be very useful in some applications. In particular, they can be used for hedging and for valuing exotic options consistently with standard ones. However, as explanations for asset pricing anomalies (such as the equity premium puzzle or the put puzzle), PE models are unsatisfactory. Suppose that one finds a PE model with multiple parameters/factors which can be calibrated reasonably well to data. Still, this does not answer many important questions: Why are certain state variables of special hedging concern to investors? Why are certain risks priced in the fist place? What are economically reasonable values for market prices of various risks?[16]

The ultimate challenge is to be able to explain historical put prices in a general equilibrium (GE) setting, which has been the main objective of our paper. Intuitively, our approach focuses on the important class of price dynamics/pricing kernels which could be rationalized by some sensible preferences in a GE context. Still, because of popularity of PE models in applied research, it might be useful to be able to test these models as well. Can our methodology help to rule out pricing kernels of PE models? We believe so.

Observe that the new restriction is rejected by the option data at a high confidence level. Intuitively, this means that a pricing kernel m_T whose projection onto v_T is only "slightly" path-dependent cannot justify the data. To formalize this intuition, we can introduce a measure $d = d(m_T)$, which for a given kernel m_T quantifies its degree of path-dependence with respect to v_T. When $d = 0$, the new restriction in (13) holds exactly. When d is nonzero, the maximum possible violation of the restriction in (13) is bounded by d. (This is somewhat in spirit of Hansen and Jagannathan (1991), who derive bounds on the pricing kernel's Sharpe ratio.) We are currently pursuing this direction of research. Preliminary results, in particular, indicate that the projected pricing kernel in the Heston model is not *sufficiently* path-dependent to rationalize historical put returns.

4.4 Implications for Option Pricing Literature

The results in Section 3.3 have important implications for the option pricing literature. In particular, the results strongly reject the specification in (11), the assumption that plays a

[16]For example, suppose that one fits the Heston model to option data. One important parameter in this model is the volatility risk premium. However, the model provides no guidance at all regarding what economically reasonable values for the volatility risk premium should be. Another common difficulty with PE models is that they often imply problematic risk preferences. For example, Chernov and Ghysels (2000, p. 414) argue that the Heston model has the counterintuitive property that, when volatility decreases to zero, the asset price risk premium increases to infinity.

central role in many theoretical and empirical papers.

Consider, for example, a recent theoretical paper by Câmara (2003). He extends the general equilibrium models in Rubinstein (1976) and Brennan (1979) to more flexible distributional assumptions. Specifically, Câmara characterizes a whole *class* of infinitely many general equilibrium models for certain families of utility functions and the joint dynamics for the aggregate consumption and the market portfolio. However, all pricing kernels in his approach come out in the form $m_T = \exp(g(v_T))$, for a general function $g(\cdot)$. This means that none of these general equilibrium models could be consistent with the results in Section 3.3.

Our results also present a challenge to the fast-growing strand of the empirical literature which uses option prices to extract the implied risk-preferences. Several recent papers, including Aït-Sahalia and Lo (2000), Jackwerth (2000), and Rosenberg and Engle (2002), develop such methodologies. These papers also present important applications. In particular, Aït-Sahalia and Lo propose to use estimated preferences in risk management by introducing the new concept of *economic* value-at-risk, or E-VaR.

While there are some variations, all methodologies essentially consist of three main steps: 1) estimate the risk-neutral density $h_t(v_T)$ from option prices, 2) estimate the objective density $f_t(v_T)$ from the underlying process, and 3) interpret the ratio $h_t(v_T)/f_t(v_T)$ as the marginal utility of the representative investor. Specifically, Aït-Sahalia and Lo (2000), Jackwerth (2000), and Bliss and Panigirtzoglou (2003) assume that the representative investor maximizes the expected value of the utility function $U(v_T)$. Under this assumption, the pricing kernel m_T satisfies (11) and $m_T = m(v_T) = h_t(v_T)/f_t(v_T)$ is equal to the marginal utility $U'(v_T)$ (times a constant). Thus, the relative risk aversion function can be computed as

$$\gamma(v_T) = -v_T \frac{U''(v_T)}{U'(v_T)} = -v_T \frac{{m_t}'(v_T)}{m_t(v_T)} = v_T \left(\frac{f_t'(v_T)}{f_t(v_T)} - \frac{h_t'(v_T)}{h_t(v_T)} \right).$$

Some anomalous findings have been reported. In particular, Jackwerth documents that, for a sizable range of wealth levels, investors seem to exhibit risk-seeking behavior, i.e., the pricing kernel $m(v_T)$ is increasing instead of decreasing and $\gamma(v_T) < 0$. Some authors refer to this finding as the "pricing kernel puzzle." The results in Section 3.3, however, suggest that the pricing kernel puzzle might be *spurious*, in the sense that it could be an artifact of the incorrect assumption. Stated differently, the pricing kernel puzzle is a puzzle *only if* the assumed specification in (11) is satisfied, which we now know is not true.

Rosenberg and Engle (2002) follow a similar empirical methodology but offer a broader interpretation. They do not rely on the specification in (11) and allow the pricing kernel $m_T = m(v_T, y_T)$ to depend on other state variables y_T. In this case, the ratio $h_t(v_T)/f_t(v_T)$ is no longer the pricing kernel, but instead is the *projected* kernel $m_t^*(v_T)$. Using the projected kernel, Rosenberg and Engle define the *projected* relative risk aversion function as[17]

$$\gamma^*(v_T) := -v_T \frac{m_t^{*\prime}(v_T)}{m_t^*(v_T)} = v_T \left(\frac{f_t'(v_T)}{f_t(v_T)} - \frac{h_t'(v_T)}{h_t(v_T)} \right).$$

They estimate two parametric specifications for the projected kernel, one of which produces the projected kernel which has similar characteristics as those reported in Jackwerth (2000). That is, there is a sizable region where the projected kernel is increasing and thus $\gamma^*(v_T)$ is negative. Does it, however, mean that investors are sometimes risk-seeking? Not necessarily.

[17] Note that, empirically, the projected and unprojected risk aversion functions are computed in exactly the same way. The only difference is their underlying theoretical interpretations.

If the projected kernel $m_t^*(v_T)$ is strongly path-dependent, which we have established in Section 3.3, then the projected risk aversion $\gamma^*(v_T)$ could be *very* different from the true risk aversion $\gamma(v_T, y_T)$. In particular, it is possible to construct an example where a) the representative investor has classical von Neumann-Morgenstern preferences with utility function $U(v_T, y_T)$, where $U_v > 0$ and $U_{vv} < 0$ for all v_T and y_T; b) for all v_T and y_T, the relative risk aversion function $\gamma(v_T, y_T)$ takes reasonable values, say, between 2 and 5; and c) the projected relative risk aversion function $\gamma^*(v_T)$ is *negative* for some values of v_T.[18]

To summarize, one important implication of our paper is that the proposed methodologies, despite their considerable popularity, might be inappropriate for estimating risk-preferences and might produce misleading results. The issue here is not so much the fact of the rejection of the specification in (10) per se. After all, *every* theoretical assumption is only an approximation. Instead, the main issue is the *extent* of the rejection. Our results suggest that the specification in (10) is grossly violated, in which case the ratio $h_t(v_T)/f_t(v_T)$ cannot be interpreted as the marginal utility for the representative investor, even approximately. This important topic warrants further investigation.

5 Conclusion

In this paper, we implement a novel methodology to test rationality of asset pricing. The main advantage of the methodology is that it requires no parametric assumptions about the unobservable pricing kernel or investors' preferences. Furthermore, it can be applied even when the sample is affected by the Peso problem and when investors' beliefs are incorrect. The methodology is based on the new rationality restriction, which states that securities prices deflated by RND evaluated at the eventual outcome must follow a martingale.

We implement the new methodology in the context of the overpriced puts puzzle. The puzzle is that historical prices of puts on the S&P 500 Index have been extremely high and incompatible with the canonical asset-pricing models. The economic impact of the put mispricing appears to be very large. Simple trading strategies that sell unhedged puts would have earned extraordinary paper profits.

To investigates whether put returns could be rationalized in a possibly nonstandard equilibrium model, we test the new rationality restriction. The required information about RND is estimated nonparametrically from prices of traded options. We find that the new restriction is strongly rejected, meaning that *no* model from a broad class of models can possibly explain the put anomaly, even when allowing for the possibility of the Peso problem and incorrect beliefs. In the light of our results, one might have to 1) develop a new kind of general equilibrium models, for which the pricing kernels is strongly path-dependent with respect to the market portfolio (such models are currently not available); 2) entertain the possibility that investors are not fully rational and that they commit systematic cognitive errors; and 3) question other standard theoretical assumptions (such as the absence of market frictions). Only future research will provide a better understanding of the put puzzle.

[18] Intuitively, when the pricing kernel m_T depends on the additional state variable y_T, the projected kernel $m_t^*(v_T)$ will be path-dependent, reflecting time variation in the additional state variable. In this case, the projected kernel, in which y_T is integrated out, might differ from the unprojected kernel quite considerably.

Appendix

A Construction of Dataset

To construct our dataset we follow the following steps:

1. For both options and futures we use settlement prices. Settlement prices (as opposed to closing prices) do not suffer from nonsynchronous/stale trading of options and the bid-ask spreads. CME calculates settlement prices simultaneously for all options, based on their last bid and ask prices. Since these prices are used to determine daily margin requirements, they are carefully scrutinized by the exchange and closely watched by traders. As a result, settlement prices are less likely to suffer from recording errors and they rarely violate basic no-arbitrage restrictions. In contrast, closing prices are generally less reliable and less complete.

2. In the dataset, we match all puts and calls by trading date t, maturity T, and strike. For each pair (t, T), we drop very low (high) strikes for which put (call) price is less than 0.1. Then we form normalized option prices as explained in Section 2.1. To convert spot prices to forward prices, we approximate the risk-free rate r_f over $[t, T]$ by the rate of Tbills.

3. Since the CME options are the American type, their prices $p_t^A(k)$ and $c_t^A(k)$ are slightly higher than prices of the corresponding European options $p_t(k)$ and $c_t(k)$. The difference, however, is very small for short maturities that we focus on. This is particularly true for OTM an ATM options.[19]

To infer prices of European options $p_t(k)$ and $c_t(k)$, we proceed as follows. First, we discard all ITM options. That is, we use put prices for $k \leq 1.00$ and call prices for $k \geq 1.00$. Prices of OTM and ATM options are both more reliable and less affected by the early exercise feature. Second, we correct American option prices $p_t^A(k)$ and $c_t^A(k)$ for the value of the early exercise feature by using Barone-Adesi and Whaley (1987) approximation.[20] Third, we compute prices of ITM options through the put-call parity relationship

$$p(k) + 1 = c(k) + k.$$

4. We check option prices for violations of the no-arbitrage restrictions. To preclude arbitrage opportunities, call and put prices must be monotonic and convex functions of the strike. In particular, the call pricing function $c_t(k)$ must satisfy

$$(a) \quad c_t(k) \geq (1-k)^+, \qquad (b) \quad -1 \leq c_t'(k) \leq 0, \qquad (c) \quad c_t''(k) \geq 0.$$

The corresponding conditions for the put pricing function $p_t(k)$ follow from put-call parity. When restrictions (a)-(c) are violated, we enforce them by running the so-called *Constrained Convex Regression* (CCR). This procedure has been proposed in Bondarenko (1997) and also implemented in Bondarenko (2000). Intuitively, CCR searches for the smallest (in the sense of least squares) perturbation of option prices that restores the no-arbitrage restrictions. For most trading days, option settlement prices already satisfy the restrictions (a)-(c). Still, CCR is a useful procedure because it allows one to identify possible recording errors or typos. We eliminate an option cross-section if CCR detects substantial arbitrage violations, that is, if square root of mean squared deviation of option prices from the closest arbitrage-free prices is more than 0.1. (This filter eliminates less than 0.5% of trading days.)

5. For each pair (t, T), we estimate RND using the *Positive Convolution Approximation* (PCA) procedure of Bondarenko (2000, 2003b). PCA is a flexible, fully nonparametric method, which produces arbitrage-free estimators, controls for overfitting in small samples, and is shown to be very accurate. For the purpose of RND estimation, we require that on date-t there are at least 8 strikes for which option prices satisfy the above filters.

[19] As shown in Whaley (1986), the early exercise premium increases with the level of the risk-free rate, volatility, time to maturity, and degree to which an option is in-the-money.

[20] It is important to point out that this correction is always substantially smaller than typical bid-ask spreads. In particular, the correction generally does not exceed 0.2% of an option price.

B Theoretical Assumptions and Proof of Proposition 1

There are two additional assumptions needed for the restriction in (8). Assumption 2, which is due to Bossaerts (2003), restricts the set of possible beliefs of investors. Specifically, suppose that beliefs are partitioned into initial beliefs (priors) and beliefs conditional on the final state x_T (likelihood functions). Then, initial beliefs can be arbitrary but that conditional beliefs must be correct.

Assumption 2 *Investors' beliefs conditional on the final state are correct. That is,*

$$E_t^{x_T}[\cdot] = E_t^{S,x_T}[\cdot], \qquad \textit{for all } x_T.$$

Assumption 2 has the following interpretation. Over time, investors gradually learn the final state x_T by observing some economic "signals". Even though investors might not know the correct distribution of final states, they nevertheless understand how the signals are generated for each x_T. Assumption 1 is satisfied trivially for the standard EMH. However, EMH requires priors to be correct as well.

Assumption 3 is a technical one. It ensures that the ratio inside the conditional expectation operator in (8) is always well-defined.

Assumption 3 *For all histories and all x_T, the risk-neutral density $h_s(x_T) > 0$.*

Proof of Proposition 1

The proof relies on the following two observations. First, the risk-neutral and subjective densities for the final state x_T are related to each other via the projected pricing kernel as follows:

$$h_t(x_T) = \frac{m_t^*(x_T)g_t(x_T)}{m_t}, \qquad h_s(x_T) = \frac{m_s^*(x_T)g_s(x_T)}{m_s}, \tag{15}$$

Second, for any random variable Y_s (measurable with respect to information at time-s):

$$E_t^{S,x}[Y_s] = \frac{E_t^S[Y_s g_s(x)]}{E_t^S[g_s(x)]} = \frac{E_t^S[Y_s g_s(x)]}{g_t(x)}. \tag{16}$$

Therefore, for any security

$$E_t^x\left[\frac{Z_s}{h_s(x)}\right] = E_t^{S,x}\left[\frac{Z_s}{h_s(x)}\right] = \frac{E_t^S\left[\frac{Z_s}{h_s(x)}g_s(x)\right]}{g_t(x)} = \frac{E_t^S\left[\frac{Z_s m_s}{m_s^*(x)}\right]}{g_t(x)} = \frac{Z_t m_t}{m_t^*(x)g_t(x)} = \frac{Z_t}{h_t(x)},$$

where we have used Assumptions 1-3, facts (15)-(16), and condition $E_t[Z_s m_s] = Z_t m_t$. □

C Parametric Example

This appendix illustrates our model-independent approach with the help of parametric example. In a tractable model, the example allows one to study the properties of the new restriction in the presence of 1) risk-aversion, 2) incorrect beliefs, and 3) selection biases. As a special case, the example produces the Black-Scholes model. (All technical details are collected in Appendix D.)

C.1 Economy

Consider a continuous-time finite-horizon economy. There is one risky asset, whose price is v_t. The risk-free rate is zero. A representative investor maximizes the expected value of the utility function $U(v_T)$. Also traded are various derivative securities in zero net supply. As a state variable, it is convenient to use log of the asset's price $x_t = \log v_t$. We assume that x_T is normally distributed. Specifically, the time-0 objective density is $f_0(x_T) = f(x_T, T; x_0, 0) = n(x_T; u_0, \eta_0^2)$, where for all μ and σ

$$n(x_T; \mu, \sigma^2) := \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{(x_T-\mu)^2}{2\sigma^2}\right].$$

Initial beliefs are represented by a normal density $g_0(x_T) = g(x_T, T; x_0, 0) = n(x_T; w_0, \sigma_0^2)$, where w_0 and σ_0^2 are the *subjective* mean and variance. In general, $(w_0, \sigma_0) \neq (u_0, \eta_0)$. The representative investor exhibits *Constant Relative Risk Aversion* (CRRA) with

$$U(v_T) = \begin{cases} \frac{1}{1-\gamma} v_T^{1-\gamma}, & \text{if } \gamma \neq 1 \\ \log v_T, & \text{if } \gamma = 1. \end{cases} \tag{17}$$

This implies that the risk-neutral density is also normal, $h_0(x_T) = h(x_T, T; x_0, 0) = n(x_T; \nu_0, \sigma_0^2)$. The risk-neutral and subjective densities have the same variances but different means, with $\nu_0 = w_0 - \gamma\sigma_0^2$. For the CRRA preferences, the pricing kernel is $m_T = v_T^{-\gamma} = e^{-\gamma x_T}$.

Over time, investors learn about the final value $x_T = x$ by observing a continuous flow of signals

$$d\Psi_t = x\,dt + \phi_t dB_t, \qquad \Psi_0 = 0, \tag{18}$$

where B_t is the standard Brownian motion. Intuitively, an incremental signal $d\Psi_t$ is normally distributed with mean xdt and variance $\phi_t^2 dt$, where ϕ_t is a given function of t. (A specific choice for ϕ_t will ensure that v_t follows the geometric Brownian motion as in the Black-Scholes model.)

Information arrival in (18) implies that, at any time $t < T$, the three probability densities are normal, $f_t(x_T) = n(x_T; u_t, \eta_t^2)$, $g_t(x_T) = n(x_T; w_t, \sigma_t^2)$ and $h_t(x_T) = n(x_T; \nu_t, \sigma_t^2)$, where

$$u_t - x = \frac{\eta_t^2}{\eta_0^2}(u_0 - x) + \eta_t^2 \int_0^t \frac{dB_s}{\phi_s}, \qquad \frac{1}{\eta_t^2} = \frac{1}{\eta_0^2} + \int_0^t \frac{ds}{\phi_s^2}.$$

$$w_t - x = \frac{\sigma_t^2}{\sigma_0^2}(w_0 - x) + \sigma_t^2 \int_0^t \frac{dB_s}{\phi_s}, \qquad \frac{1}{\sigma_t^2} = \frac{1}{\sigma_0^2} + \int_0^t \frac{ds}{\phi_s^2},$$

and $\nu_t = w_t - \gamma\sigma_t^2$. The stochastic processes for u_t and ν_t are[21]

$$du_t = \frac{\eta_t^2}{\phi_t^2}(x - u_t)dt + \frac{\eta_t^2}{\phi_t} dB_t, \qquad d\nu_t = \frac{\sigma_t^2}{\phi_t^2}(x - \nu_t)dt + \frac{\sigma_t^2}{\phi_t} dB_t.$$

Because of normality, u_t, w_t, and ν_t can be interpreted as the time-t objective, subjective, and risk-neutral expectations of the final outcome x. Since $x_t = (\nu_t + 0.5\sigma_t^2)$ and $v_t = e^{x_t}$, it follows that the instantaneous return on the asset's price is $dv_t/v_t = d\nu_t$.

C.2 Special Case

Consider first a special case where $\sigma_0 = \sigma\sqrt{T}$ and $\phi_t = \sigma(T - t)$ for some constant σ, implying that $\sigma_t^2 = \sigma^2(T - t)$. Conditionally on $x_T = x$, the process for x_t is a (generalized) Brownian Bridge:

$$dx_t = \frac{x - x_t}{T - t} dt + \sigma dB_t.$$

With *no* conditioning on the final outcome, the process for the asset's price v_t can be derived as

$$\frac{dv_t}{v_t} = \frac{1}{T - t}(u_t - \nu_t)dt + \sigma dB_t.$$

In general, the drift in the above formula depends on initial beliefs. Suppose that $(u_0, \eta_0) = (\nu_0 + (\Delta + \gamma)\sigma_0^2, \sigma_0)$ for some Δ. Under this specification of beliefs, the standard deviation is unbiased, while the bias in the mean is $\Delta\sigma_0^2$. The process for v_t then reduces to

$$\frac{dv_t}{v_t} = (\Delta + \gamma)\sigma^2 dt + \sigma dB_t.$$

[21] By assuming that ϕ_t satisfies $\lim_{t\to T} \int_0^t \frac{ds}{\phi_s^2} = \infty$, we ensure that the three densities all converge to the delta function $\delta(x)$ at time T, in the sense that

$$\lim_{t\to T} \eta_t = \lim_{t\to T} \sigma_t = 0, \qquad \text{and} \qquad \lim_{t\to T} u_t = \lim_{t\to T} w_t = \lim_{t\to T} \nu_t = x.$$

(Clearly, the classical Black-Scholes model corresponds to the case of correct beliefs with $\Delta = 0$.)

Remark 1. The above equation has an interesting implication. Consider an empiricist who only observes the objective process. The empiricist is unable to disentangle the effects of risk aversion and biased beliefs, because the empiricist can only observe the sum $(\Delta + \gamma)$, but not γ and Δ separately. This means that the same securities prices can result from either risk-aversion or biased beliefs, or some combination of the two. In particular, the same risk premium $(\gamma + \Delta)\sigma$ can arise in any economy for which risk-aversion γ' and the bias in beliefs Δ' are such that $\gamma' + \Delta' = \gamma + \Delta$. (Bondarenko (2003a) discusses general conditions under which preferences and beliefs are observationally equivalent.)

C.3 Comparison of Rationality Restrictions

We now contrast the two rationality restrictions, the standard one in (6) and the new one in (9). Although these restrictions apply to all securities, we only illustrate them for standard puts.

Suppose that the period $[0, s]$ is fixed, where $0 < s < T$. Consider the put option on the asset with moneyness $k = K/v_0$ and maturity s. The normalized put price is $p_0(k)$, and the put net return over $[0, s]$ is $r_p(K)$. Let $\lambda^x = h_0(x)/h_s(x)$ be the inverse of the return of RND evaluated at the final outcome x, and let $m = m_s/m_t$. Consider the following three moments (derived in Appendix D):

$$\begin{aligned} I_1 &:= E_0[r_p(k)] = \frac{p^{BS}(k; \nu^*_{0s}, \sigma^{*2}_{0s})}{p^{BS}(k; \nu_{0s}, \sigma^2_{0s})} - 1, \\ I_2 &:= E_0[m r_p(k)] = \left(\frac{p^{BS}(k; \nu^*_{0s} - \gamma\sigma^{*2}_{0s}, \sigma^{*2}_{0s})}{p^{BS}(k; \nu_{0s}, \sigma^2_{0s})} - 1 \right) \cdot e^D, \\ I_3^x &:= E_0^x[\lambda^x r_p(k)] = 0, \end{aligned}$$

where

$$\nu_{0s} := -0.5(\sigma_0^2 - \sigma_s^2), \qquad \nu^*_{0s} := -0.5(\sigma_0^2 - \sigma_s^2) + \left(1 - \frac{\sigma_s^2}{\sigma_0^2}\right)(u_0 - \nu_0),$$

$$\sigma^2_{0s} := \sigma_0^2 - \sigma_s^2, \qquad \sigma^{*2}_{0s} := (\sigma_0^2 - \sigma_s^2) + \left(1 - \frac{\sigma_s^2}{\sigma_0^2}\right)^2 (\eta_0^2 - \sigma_0^2),$$

$$D := \gamma(\nu_{0s} - \nu^*_{0s}) + 0.5\gamma^2(\sigma^2_{0s} + \sigma^{*2}_{0s}).$$

The moments I_1 and I_2 are expressed in terms of the Black-Scholes price $p^{BS}(k; \mu, \sigma^2)$, where

$$p^{BS}(k; \mu, \sigma^2) := \int_{-\infty}^{\infty} (k - e^y)^+ \, n(y; \mu, \sigma^2) \, dy = k \, N\left(\frac{\ln k - \mu}{\sigma}\right) - e^{\mu + 0.5\sigma^2} \, N\left(\frac{\ln k - \mu - \sigma^2}{\sigma}\right),$$

and $N(\cdot)$ is the standard normal cdf.

In general, the moments I_1 and I_2 are different from zero. The only situation when $I_2 = 0$ is when $\sigma^{*2}_{0s} = \sigma^2_{0s}$ and $\nu^*_{0s} - \gamma\sigma^{*2}_{0s} = \nu_{0s}$. This occurs when beliefs are correct, or $(u_0, \eta_0) = (w_0, \sigma_0)$. The only situation when $I_1 = 0$ is when $\sigma^{*2}_{0s} = \sigma^2_{0s}$ and $\nu^*_{0s} = \nu_{0s}$. This occurs when $(u_0, \eta_0) = (\nu_0, \sigma_0)$, that is, when the subjective standard deviation σ_0 is unbiased while the upward bias in the subjective mean w_0 is exactly offset by risk-aversion, so that $w_0 - \gamma\sigma_0^2 = \nu_0 = u_0$. In contrast, the condition $I_3^x = 0$ holds for arbitrary beliefs and risk-aversion.

Figure 6 plots the three moments I_1, I_2, and I_3^x across strikes K for several special cases. In all cases, we assume $\sigma_0 = \sigma\sqrt{T}$ and $\phi_t = \sigma(T - t)$ for constant σ. We set $s = 1$, $T = 2$, $\sigma = 0.04$, and $\gamma = 4$, where all parameters correspond to the monthly frequency. To allow for biased beliefs, we represent the objective mean as $u_0 = \nu_0 + (\Delta + \gamma)\sigma_0^2$. The first four cases in Figure 6 are

I. Correct beliefs, $\Delta = 0$, $\eta_0 = \sigma_0$ (this is the classical Black-Scholes model);

II. Incorrect beliefs with a biased mean, $\Delta = 2$, $\eta_0 = \sigma_0$;

III. Incorrect beliefs with a biased standard deviation, $\Delta = 0$, $\eta_0 = 0.85\sigma_0$;

IV. Incorrect beliefs with biased mean and standard deviation, $\Delta = 2$, $\eta_0 = 0.85\sigma_0$.

From Figure 6, the moment I_3^x is equal to zero for all moneyness k and for all cases. The moment I_2 is zero only when beliefs are correct (case I). Recall that, to test the restriction $I_2 = 0$, one must specify the pricing kernel, which is unobservable in practice. Even when the true pricing kernel is available, the moment I_2 will still be nonzero if beliefs are incorrect. In particular, I_2 is negative for all k when investors underestimate the objective mean u_0 (case II), overestimate the objective standard deviation (case III), or both (case IV). The moment I_1 corresponds to the sample average return and is affected by both risk-aversion and biases in beliefs. For cases I-IV, I_1 is negative for all strikes.

Remark 2. In our analysis, we rely on the CRRA-lognormal setup only because of analytical tractability. However, the condition $I_3^x = 0$ will continue to hold for general utility functions $U(v_T)$, initial beliefs, and information flow in (18).

C.4 Selection Bias

We now illustrate the new restriction in (9) in the presence of a selection bias. Let A denote a subset of final states and suppose that the empiricist has a sample of histories with $x_T \in A$. For example, if $A = \{x_T \,|\, x_T \geq x^c\}$, then the sample only includes returns above the critical value x^c. Such a selection bias may contribute to the apparent put overpricing. Very low asset's returns happen infrequently and may not be observed in a small sample. Nevertheless, since low asset's returns correspond to very high returns of OTM puts, it is possible that the omission of these observations might significantly distort the small-sample moments.

In the presence of the selection bias, the empiricist considers the following moments:

$$\begin{aligned}\bar{I}_1 = \bar{I}_1(A) &:= E_0[r_p(k) \,|\, x_T \in A],\\ \bar{I}_2 = \bar{I}_2(A) &:= E_0[mr_p(k) \,|\, x_T \in A],\\ \bar{I}_3 = \bar{I}_3(A) &:= E_0[I_3^{x_T} \,|\, x_T \in A].\end{aligned}$$

These moments can be computed as explained in Appendix D. To introduce a selection bias, we choose the cutoff value x^c such that 2% of the asset's lowest returns are discarded. We plot $\bar{I}_1$, $\bar{I}_2$, and $\bar{I}_3$ in the bottom panels of Figure 6, for two cases:

V. Selection bias, correct beliefs;

VI. Selection bias, incorrect beliefs with biased mean and standard deviation, $\Delta = 2$, $\eta_0 = 0.85\sigma_0$.

Figure 6 shows that omitting a small number of extreme returns can generate a sizable (but spurious) mispricing of puts, especially for deep OTM puts. The figure also demonstrates that selection bias does not affect the condition $\bar{I}_3 = 0$.

Remark 3. The condition $\bar{I}_3 = 0$ will continue to hold for many other subsets A, for example, $A = \{x_T \,|\, x_l \leq x_T \leq x_h\}$ for $x_l < x_h$. More importantly, *ex post* frequencies of final states x_T can deviate from *ex ante* probabilities $f_0(x_T)$ in an arbitrary way and, still, the restriction in (9) must hold.

D Technical Details for Parametric Example

This appendix provides technical details for the parametric example in Appendix C. In derivations, the following basic properties of normal densities are used repeatedly.

$$\text{P1:} \qquad \int_{-\infty}^{\infty} n(x;\mu,\sigma^2)\, e^{\alpha x}\, dx = e^{0.5\alpha^2\sigma^2 + \alpha\mu}.$$

$$\text{P2:} \qquad n(x;y,\eta^2)\, n(y;\mu,\sigma^2) = n(x;\mu,\sigma^2+\eta^2)\, n\left(y; \frac{\frac{\mu}{\sigma^2}+\frac{x}{\eta^2}}{\frac{1}{\sigma^2}+\frac{1}{\eta^2}}, \frac{1}{\frac{1}{\sigma^2}+\frac{1}{\eta^2}}\right).$$

$$\text{P3:} \qquad \int_{-\infty}^{\infty} n(x;\alpha+\beta y,\eta^2)\, n(y;\mu,\sigma^2)\, dy = n(x;\alpha+\beta\mu,\beta^2\sigma^2+\eta^2).$$

For example, the pricing kernel m_t can be derived from property P1 as

$$m_t = E_t^S[m_T] = \int_{-\infty}^{\infty} e^{-\gamma x_T}\, g_t(x_T)\, dx_T = \int_{-\infty}^{\infty} e^{-\gamma x_T}\, n(x_T; w_t, \sigma_t^2)\, dx_T = e^{-\gamma \nu_t - 0.5\gamma^2\sigma_t^2}.$$

D.1 Derivation of Moments I_1, I_2, and I_3^x

We first derive the objective density $f(x_s, s; x_0, 0)$. Let $f(x_s, s; x_0, 0, x_T, T)$ denote the objective density of x_s conditional on both x_0 and x_T. From

$$x_s = \sigma_s^2 \left(\frac{x_0}{\sigma_0^2} + \int_0^s \frac{d\Psi_\tau}{\phi_\tau^2} \right) = \sigma_s^2 \left(\frac{x_0}{\sigma_0^2} + x_T \left(\frac{1}{\sigma_s^2} - \frac{1}{\sigma_0^2} \right) + \int_0^s \frac{dB_\tau}{\phi_\tau} \right),$$

we obtain that

$$\begin{aligned}
f(x_s, s; x_0, 0, x_T, T) &= n\left(x_s; x_0 \frac{\sigma_s^2}{\sigma_0^2} + x_T \left(1 - \frac{\sigma_s^2}{\sigma_0^2}\right), \sigma_s^2 \left(1 - \frac{\sigma_s^2}{\sigma_0^2}\right) \right), \\
f(x_s, s; x_0, 0) &= \int_{-\infty}^{\infty} f(x_s, s; x_0, 0, x_T, T)\, f_0(x_T)\, dx_T \\
&= \int_{-\infty}^{\infty} n\left(x_s; x_0 \frac{\sigma_s^2}{\sigma_0^2} + x_T \left(1 - \frac{\sigma_s^2}{\sigma_0^2}\right), \sigma_s^2 \left(1 - \frac{\sigma_s^2}{\sigma_0^2}\right) \right) n(x_T; u_0, \eta_0^2)\, dx_T \\
&= n\left(x_s; x_0 \frac{\sigma_s^2}{\sigma_0^2} + u_0 \left(1 - \frac{\sigma_s^2}{\sigma_0^2}\right), \sigma_s^2 \left(1 - \frac{\sigma_s^2}{\sigma_0^2}\right) + \eta_0^2 \left(1 - \frac{\sigma_s^2}{\sigma_0^2}\right)^2 \right) \\
&= n(x_s; \nu_{0s}^* + x_0, \sigma_{0s}^{*2}).
\end{aligned}$$

Therefore,

$$E_0[p_s(k)] = \int_{-\infty}^{\infty} \left(k - e^{x_s - x_0}\right)^+ f(x_s, s; x_0, 0)\, dx_s$$

$$= \int_{-\infty}^{\infty} \left(k - e^{x_s - x_0}\right)^+ n(x_s; \nu_{0s}^* + x_0, \sigma_{0s}^{*2})\, dx_s = p^{BS}(k; \nu_{0s}^*, \sigma_{0s}^{*2}).$$

Similarly, the risk-neutral density $h(x_s, s; x_0, 0) = n(x_s; \nu_{0s} + x_0, \sigma_{0s}^2)$, and the normalized price

$$p_0(k) = \int_{-\infty}^{\infty} (k - e^{x_s - x_0})^+\, h(x_s, s; x_0, 0)\, dx_s = p^{BS}(k; \nu_{0s}, \sigma_{0s}^2).$$

Recall now that $m = m_s/m_0 = e^{-\gamma(\nu_s - \nu_0) - 0.5\gamma^2(\sigma_s^2 - \sigma_0^2)}$ and $x_s = \nu_s + 0.5\sigma_s^2$. Therefore,

$$\begin{aligned}
E_0[m] &= \int_{-\infty}^{\infty} e^{-\gamma(\nu_s - \nu_0) - 0.5\gamma^2(\sigma_s^2 - \sigma_0^2)}\, f(x_s, s; x_0, 0)\, dx_s \\
&= e^D \int_{-\infty}^{\infty} e^{-\gamma(x_s - \nu_{0s}^* - x_0) - 0.5\gamma^2\sigma_{0s}^{*2}}\, n(x_s; \nu_{0s}^* + x_0, \sigma_{0s}^{*2})\, dx_s \\
&= e^D \int_{-\infty}^{\infty} n(x_s; \nu_{0s}^* + x_0 - \gamma\sigma_{0s}^{*2}, \sigma_{0s}^{*2})\, dx_s = e^D, \\
E_0[m p_s(k)] &= \int_{-\infty}^{\infty} p_s(k)\, e^{-\gamma(\nu_s - \nu_0) - 0.5\gamma^2(\sigma_s^2 - \sigma_0^2)}\, f(x_s, s; x_0, 0)\, dx_s \\
&= e^D \int_{-\infty}^{\infty} \left(k - e^{x_s - x_0}\right)^+ n(x_s; \nu_{0s}^* + x_0 - \gamma\sigma_{0s}^{*2}, \sigma_{0s}^{*2})\, dx_s \\
&= e^D\, p^{BS}(k; \nu_{0s}^* - \gamma\sigma_{0s}^{*2}, \sigma_{0s}^{*2}).
\end{aligned}$$

Finally, the condition $I_3^x = 0$ follows from the general result in Bondarenko (2003a). Alternatively, it can be verified directly as follows. By property P2,

$$h_s(x)\, h(x_s, s; x_0, 0) = n(x; \nu_s, \sigma_s^2)\, n(x_s; \nu_0 + 0.5\sigma_s^2, \sigma_0^2 - \sigma_s^2) = n(x; \nu_s, \sigma_s^2)\, n(\nu_s; \nu_0, \sigma_0^2 - \sigma_s^2)$$

$$= n(x;\nu_0,\sigma_0^2)\, n\left(\nu_s;\nu_0\frac{\sigma_s^2}{\sigma_0^2}+x\left(1-\frac{\sigma_s^2}{\sigma_0^2}\right),\sigma_s^2\left(1-\frac{\sigma_s^2}{\sigma_0^2}\right)\right) = h_0(x)\, f(x_s,s;x_0,0,x,T).$$

Therefore,

$$E_0^x\left[\lambda^x\, p_s(k)\right] = \int_{-\infty}^{\infty} p_s(k)\,\frac{h_0(x)}{h_s(x)}\, f(x_s,s;x_0,0,x,T)\,dx_s = \int_{-\infty}^{\infty} p_s(k)\, h(x_s,s;x_0,0)\,dx_s = p_0(k).$$

D.2 Derivation of Moments $\bar{I}_1$, $\bar{I}_2$, and $\bar{I}_3$

The moments $\bar{I}_1$ and $\bar{I}_2$ can be computed by numerical integration. The following approach allows us to reduce a two-dimensional integration to a one-dimensional one. We assume that $A = \{x_T \,|\, x_l \le x_T \le x_h\}$ for some $x_l < x_h$ and define

$$\begin{aligned}
G(\nu_s) &:= \frac{\int_{x_l}^{x_h} f_0(x_T) f(x_s,s;x_0,0,x_T,T)\,dx_T}{\int_{x_l}^{x_h} f_0(x_T)\,dx_T} \\
&= \frac{\int_{x_l}^{x_h} n(x_T;u_0,\eta_0^2)\, n\left(x_s;x_0\frac{\sigma_s^2}{\sigma_0^2}+x_T\left(1-\frac{\sigma_s^2}{\sigma_0^2}\right),\sigma_s^2\left(1-\frac{\sigma_s^2}{\sigma_0^2}\right)\right)\,dx_T}{\int_{x_l}^{x_h} n(x_T;u_0,\eta_0^2)\,dx_T} \\
&= n\left(x_s;\frac{\sigma_s^2}{\sigma_0^2}\nu_0+\left(1-\frac{\sigma_s^2}{\sigma_0^2}\right)u_0+0.5\sigma_s^2,\sigma_{0s}^{*2}\right)\frac{\int_{x_l}^{x_h} n(x_T;\Lambda,\Sigma^2)\,dx_T}{\int_{x_l}^{x_h} n(x_T;u_0,\eta_0^2)\,dx_T} \\
&= n\left(\nu_s;\frac{\sigma_s^2}{\sigma_0^2}\nu_0+\left(1-\frac{\sigma_s^2}{\sigma_0^2}\right)u_0,\sigma_{0s}^{*2}\right)\frac{N\left(\frac{x_h-\Lambda}{\Sigma}\right)-N\left(\frac{x_l-\Lambda}{\Sigma}\right)}{N\left(\frac{x_h-u_0}{\eta_0}\right)-N\left(\frac{x_l-u_0}{\eta_0}\right)},
\end{aligned}$$

where

$$\Lambda = \left(\frac{u_0}{\eta_0^2}+\frac{\nu_s}{\sigma_s^2}-\frac{\nu_0}{\sigma_0^2}\right)\Sigma^2, \qquad \frac{1}{\Sigma^2} = \frac{1}{\eta_0^2}+\frac{1}{\sigma_s^2}-\frac{1}{\sigma_0^2}.$$

Using the expression for $G(\nu_s)$, we can now integrate numerically the following expectations:

$$\begin{aligned}
E_0[p_s(k)\,|\,x_T\in A] &= \int_{-\infty}^{\infty}\left(k-e^{\nu_s+0.5\sigma_s^2-x_0}\right)^+ G(\nu_s)\,d\nu_s, \\
E_0[m\,|\,x_T\in A] &= \int_{-\infty}^{\infty} e^{-\gamma(\nu_s-\nu_0)-0.5\gamma^2(\sigma_s^2-\sigma_0^2)}\, G(\nu_s)\,d\nu_s, \\
E_0[mp_s(k)\,|\,x_T\in A] &= \int_{-\infty}^{\infty} e^{-\gamma(\nu_s-\nu_0)-0.5\gamma^2(\sigma_s^2-\sigma_0^2)}\left(k-e^{\nu_s+0.5\sigma_s^2-x_0}\right)^+ G(\nu_s)\,d\nu_s.
\end{aligned}$$

Finally, since $I_3^{x_T}=0$ for all x_T, we immediately obtain that $\bar{I}_3 = E_0[I_3^{x_T}\,|\,x_T\in A] = 0$.

References

Aït-Sahalia, Y., and A. Lo, 2000, "Nonparametric Risk Management and Implied Risk Aversion," *Journal of Econometrics* 94, 9-51.

Aït-Sahalia, Y., Y. Wang, and F. Yared, 2001, "Do Options Markets Correctly Price the Probabilities of Movement of the Underlying Asset?" *Journal of Econometrics* 102, 67-110.

Bakshi, G., and N. Kapadia, 2003, "Delta-Hedged Gains and the Negative Market Volatility Premium," *Review of Financial Studies* 16, 527-566.

Banz, R., and M. Miller, 1978, "Prices for State-Contingent Claims: Some Estimates and Applications," *Journal of Business* 51, 653-672.

Bekaert, G., R. Hodrick, and D. Marshall, 1995, "Peso Problem' Explanations For Term Structure Anomalies," Stanford University working paper.

Barone-Adesi, G., and Whaley, R., 1987, "Efficient Analytical Approximation of American Option Values," *Journal of Finance* 41, 301-320.

Bates, D., 2000, "Post-'87 Crash Fears in the S&P 500 Futures Option Market," *Journal of Econometrics* 94, 181-238.

Bick, A., 1990, "On Viable Diffusion Price Processes of the Market Portfolio," *Journal of Finance* 45, 673-689.

Bliss, R., and N. Panigirtzoglou, 2003, "Option-Implied Risk Aversion Estimates," *Journal of Finance*, forthcoming.

Bollen, N., and R. Whaley, 2003, "What Determines the Shape of Implied Volatility Functions," *Journal of Finance*, forthcoming.

Bondarenko, O., 1997, "Testing Rationality of Financial Markets: An Application to S&P 500 Index Options," Caltech working paper.

Bondarenko, O., 2000, "Recovering Risk-Neutral Densities: A New Nonparametric Approach," UIC working paper.

Bondarenko, O., 2003a, "Statistical Arbitrage and Securities Prices," *Review of Financial Studies* 16, 875-919.

Bondarenko, O., 2003b, "Estimation of Risk-Neutral Densities Using Positive Convolution Approximation," *Journal of Econometrics* 116, 85-112.

Bondarenko, O., and P. Bossaerts, 2000, "Expectations and Learning in Iowa," *Journal of Banking and Finance* 24, 1535-1555.

Bossaerts, P., 2003, "Filtering Returns For Unspecified Biases In Priors When Testing Asset Pricing Theory" *Review of Economic Studies* 70, 1-24.

Bossaerts, P., 1999, "Learning-Induced Securities Price Volatility," Caltech working paper.

Bossaerts, P., and P. Hillion, 2001, "IPO Post-Issue Markets: Questionable Predilections But Diligent Learners?" *Review of Economics and Statistics* 83 (2), 1-15.

Breeden, D., and R. Litzenberger, 1978, "Prices of State Contingent Claims Implicit in Options Prices," *Journal of Business* 51, 621-652.

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

Câmara, A., 2003, "A Generalization of the Brennan-Rubinstein Approach for the Pricing of Derivatives," *Journal of Finance* 58, 805-819.

Chernov, M., and E. Ghysels, 2000, "A Study Towards a Unified Approach to the Joint Estimation of Objective and Risk Neutral Measures for the Purpose of Options Valuation," *Journal of Financial Economics* 56, 407-458.

Cochrane, J., 2001, *Asset Pricing,* Princeton University Press, Princeton, NJ.

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

Dumas, B., J. Fleming, and R. Whaley, 1998, "Implied Volatility Functions: Empirical Tests," *Journal of Finance* 53, 2059-2106.

Goetzmann, W., J. Ingersoll, M. Spiegel, and I. Welch, 2002, "Sharpening Sharpe Ratios," working paper, Yale School of Management.

Hansen, L., and R. Jagannathan, 1991, "Implications of Security Market Data for Models of Dynamic Economies," *Journal of Political Economy* 99, 225-262.

Hansen, L., and R. Richard, 1987, "The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models," *Econometrica* 55, 587-613.

Harrison, J., and D. Kreps, 1979, "Martingales and Arbitrage in Multiperiod Securities Markets," *Journal of Economic Theory* 20, 381-408.

Harrison, M., and S. Pliska, 1981, "Martingales and Stochastic Integrals in the Theory of Continuous Trading," *Stochastic Process and Their Applications* 11, 215-260.

He, H., and H. Leland, 1993, "On Equilibrium Asset Price Processes," *Review of Financial Studies* 6, 593-617.

Heston, S., 1993, "A Closed-form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options," *Review of Financial Studies* 6, 327-343.

Hull, J., and A. White, 1987, "The Pricing of Options on Assets with Stochastic Volatilities," *Journal of Finance* 42, 281-300.

Jackwerth, J., 1999, "Option Implied Risk-Neutral Distributions and Implied Binomial Trees: A Literature Review," *Journal of Derivatives* 7, 66-82.

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

Merton, R., 1980, "On Estimating the Expected Return on the Market: An Exploratory Investigation," *Journal of Financial Economics* 8, 323-361.

Modigliani, F., and L. Modigliani, 1997, "Risk-Adjusted Performance," *Journal of Portfolio Management* 23 (2), 45-54.

Naik, V., and M. Lee, 1990, "General Equilibrium Pricing of Options on the Market Portfolio with Discontinuous Returns," *Review of Financial Studies* 3, 493-521.

Rosenberg, J., and R. Engle, 2002, "Empirical pricing kernels," *Journal of Financial Economics* 64, 341-372

Ross, S., 1976, "Options and Efficiency," *Quarterly Journal of Economics* 90, 75-89.

Rubinstein, M., 1976, "The Valuation of Uncertain Income Streams and the Pricing of Options," *Bell Journal of Economics* 7, 407-425.

Whaley, R., 1986, "Valuation of American Futures Options: Theory and Empirical Tests," *Journal of Finance* 41, 127-150.

Table 1: **Monthly Option Returns**

I. Put return $r_p(k)$ for different k

	0.94	0.96	0.98	1.00	1.02	1.04	1.06
n	67	109	159	161	161	161	150
Mean	-0.95	-0.58	-0.54	-0.39	-0.26	-0.17	-0.11
1%	-1.00	-0.87	-0.77	-0.59	-0.42	-0.30	-0.22
5%	-0.99	-0.80	-0.72	-0.54	-0.37	-0.28	-0.19
95%	-0.89	-0.35	-0.36	-0.24	-0.14	-0.07	-0.03
99%	-0.87	-0.22	-0.26	-0.18	-0.08	-0.03	0.00
Min.	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
Med.	-1.00	-1.00	-1.00	-1.00	-0.58	-0.32	-0.15
Max.	0.42	9.53	7.25	5.40	3.94	2.35	1.90

II. Call return $r_c(k)$ for different k

	0.94	0.96	0.98	1.00	1.02	1.04	1.06
n	160	161	161	161	143	69	21
Mean	0.06	0.06	0.05	0.04	-0.04	-0.06	0.21
1%	-0.03	-0.05	-0.11	-0.18	-0.36	-0.62	-0.83
5%	-0.01	-0.02	-0.07	-0.12	-0.27	-0.44	-0.65
95%	0.13	0.15	0.17	0.20	0.18	0.44	1.26
99%	0.15	0.19	0.20	0.28	0.27	0.63	1.82
Min.	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
Med.	0.03	-0.01	-0.12	-0.57	-1.00	-1.00	-1.00
Max.	1.63	2.07	2.49	4.73	7.35	8.36	10.96

Notes: Sample Period is 08/87–12/00. Statistics are reported for different strike-to-underlying ratio $k = K/v_t$; n is the number of observations. The confidence intervals (1%, 5%, 95%, and 99%) are constructed using a bootstrap with 1000 resamples.

Table 2: **Risk Characteristics of Monthly Returns, Puts and S&P 500 Futures**

	Put, k							S&P
	0.94	0.96	0.98	1.00	1.02	1.04	1.06	500
n	67	109	159	161	161	161	150	161
Mean	-0.95	-0.58	-0.54	-0.39	-0.26	-0.17	-0.12	0.007
Std. dev.	0.24	1.51	1.36	1.13	0.93	0.73	0.60	0.040
Skewness	4.90	4.56	3.85	2.52	1.64	1.03	0.78	-0.34
β	-2.04	-20.42	-23.07	-22.78	-20.75	-17.36	-14.34	1.00
α	-0.94*	-0.43*	-0.38*	-0.23*	-0.11	-0.05	-0.01	0.00
SR	-3.93	-0.38	-0.40	-0.35	-0.28	-0.24	-0.18	0.18
TM	0.46	0.021	0.016	0.010	0.005	0.003	0.001	0.00
M^2	-0.16	-0.015	-0.016	-0.014	-0.011	-0.009	-0.007	0.007
Implied γ	130.9	9.0	9.7	8.3	6.7	5.7	4.3	4.3

Notes: Sample Period is 08/87–12/00. Statistics are reported for monthly returns of puts with different k and the underlying S&P 500 futures. Net returns are in excess of the risk-free rate and computed over 1 month period prior to the maturity date; n is the number of observations. SR is the Sharpe ratio, TM is the Treynor's measure, M^2 is M-squared. For Jensen's α, asterisk (*) denotes significance at the 1% level. "Implied γ" is the coefficient of relative risk aversion in Rubinstein (1976) model that is required to justify realized returns.

Table 3: **Highest Monthly Put Returns**

	Put, k							S&P	ATM
Holding period	0.94	0.96	0.98	1.00	1.02	1.04	1.06	500	Vol.
1987: 09/18-10/16	n/a	9.53	7.02	3.88	2.99	2.35	1.90	-0.11	0.19
1987: 10/16-11/20	n/a	5.21	4.27	3.46	2.44	2.00	1.69	-0.14	0.28
1990: 07/20-08/17	n/a	n/a	7.25	5.40	3.94	2.22	1.75	-0.10	0.14
1994: 03/18-04/15	n/a	n/a	3.94	3.17	1.98	1.20	0.98	-0.05	0.11
1998: 07/17-08/21	n/a	6.39	5.46	4.03	2.78	2.06	1.56	-0.09	0.14

Notes: Table reports on five holding periods that correspond to the highest put returns in the sample. Sample Period is 08/87–12/00. Also reported is return on the underlying S&P 500 futures. Returns are in excess of the risk-free rate. The last column is (annualized) one-month ATM implied volatility, computed at the beginning of each holding period.

Table 4: **The number of crashes needed for AR=0**

	k						
	0.94	0.96	0.98	1.00	1.02	1.04	1.06
Oct 87 return	n/a	12	20	18	17	14	10
per year	n/a	1.30	1.43	1.29	1.21	1.02	0.81
Highest return	n/a	7	12	12	10	12	9
per year	n/a	0.79	0.91	0.89	0.80	0.90	0.74

Notes: Sample Period is 08/87–12/00. For each k, the table reports how many artificial extreme returns (October 1987 or the largest ever) are required to reconcile the put anomaly. Also shown is the corresponding number of crashes per year. See Section 2.3.2 for details.

Table 5: **Drift of the S&P 500 Index needed for AR=0**

	k						
	0.94	0.96	0.98	1.00	1.02	1.04	1.06
Negative drift	0.047	0.027	0.020	0.015	0.011	0.009	0.007
S&P final value	0.5	13.3	47.8	111.9	221.0	285.3	401.0

Notes: Sample Period is 08/87–12/00. For each k, the table reports the value of hypothetical negative drift η that is required to reconcile the put anomaly. Also shown what the corresponding value of the S&P 500 Index would have been at the end of 2000. The drift is reported in monthly decimal terms. See Section 2.3.3 for details.

Table 6: **Put Weighted Return $\lambda r_p(k)$ for Different k**

	0.94	0.96	0.98	1.00	1.02	1.04	1.06
n	55	93	142	144	144	144	132
Mean	-0.79	-0.59	-0.51	-0.35	-0.21	-0.13	-0.08
1%	-1.06	-0.89	-0.75	-0.59	-0.42	-0.30	-0.24
5%	-0.99	-0.82	-0.70	-0.53	-0.36	-0.27	-0.20
95%	-0.63	-0.31	-0.27	-0.12	-0.02	0.02	0.05
99%	-0.57	-0.20	-0.14	-0.02	0.06	0.08	0.14
Min.	-4.35	-4.35	-4.35	-4.35	-4.35	-4.35	-4.35
Med.	-0.67	-0.68	-0.67	-0.59	-0.39	-0.21	-0.12
Max.	1.72	11.46	15.07	14.44	11.94	9.59	7.23

Notes: Sample Period is 08/87–12/00. Statistics for monthly weighted return $\lambda r_p(k)$ are reported for one-month puts with different strike-to-underlying ratio $k = K/v_t$; n is the number of observations. Return $r_p(k)$ is in excess of the risk-free rate and is computed over one-month holding periods $[t, s]$ prior to the option maturity date s. The weight $\lambda = h_t(v_T)/h_s(v_T)$ is the inverse of the return on RND evaluated at the realized future value of the underlying v_T, where T-s is one month. The confidence intervals (1%, 5%, 95%, and 99%) are constructed using a bootstrap with 1000 resamples.

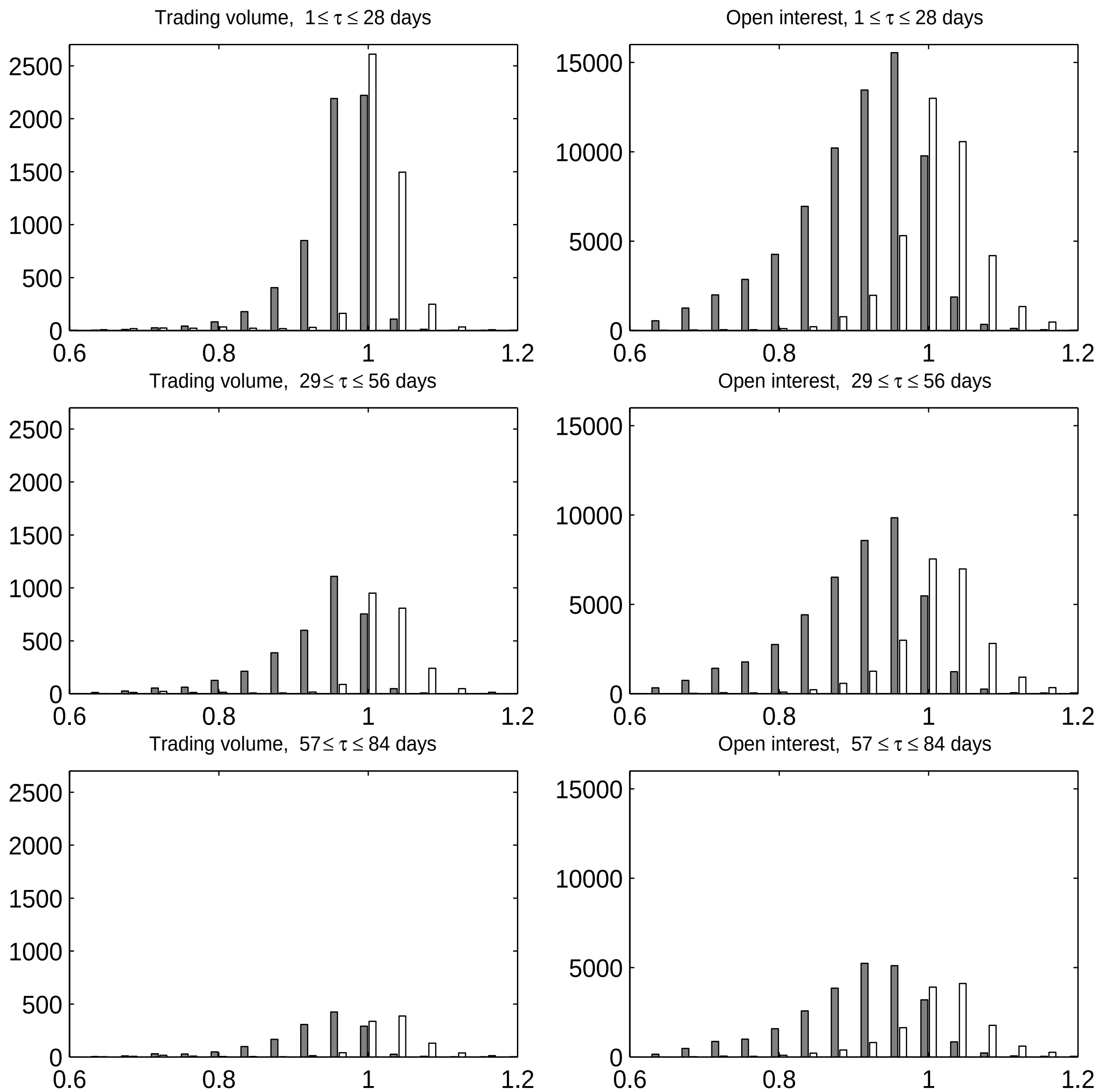

Figure 1: Average daily trading volume and open interest for puts (dark bars) and calls (white bars). The sample period is 08/87–12/00. The average statistics are computed for different strike-to-underlying ratio $k = K/v_t$, when the number of days to maturity τ is 1 to 28 (total of 3,103 trading days), 29 to 56 (total of 3,137 trading days), and 57 to 84 (total of 3,109 trading days), respectively. In November 1997, the contract multiplier for the CME options was reduced from 500 to 250. Therefore, to calculate average trading volume and open interest, the number of contracts before November 1997 is multiplied by 2.

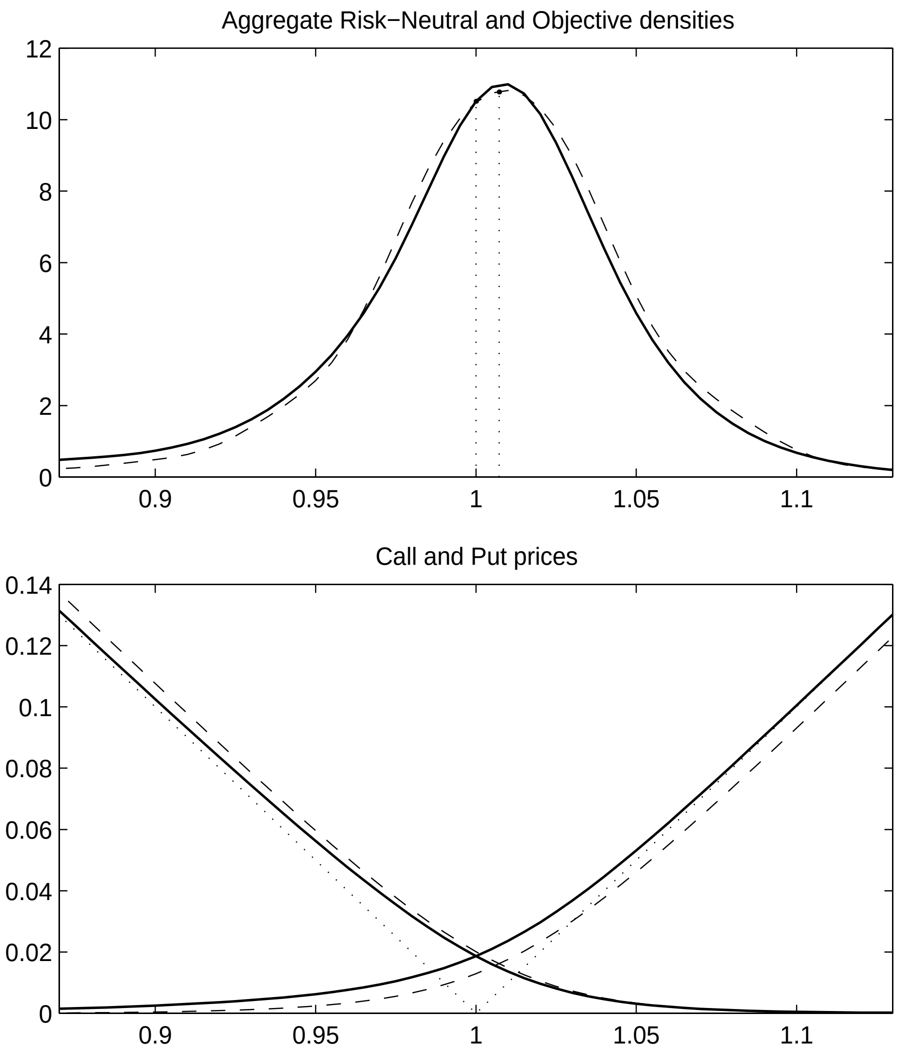

Figure 2: The top panel plots the aggregate risk-neutral density (ARND) and the objective density (OD) versus the strike-to-underlying ratio k, when time to maturity is 1 month. ARND (the solid line) is the pointwise average of 161 individual RNDs. OD (the dashed line) is estimated using the kernel method. The dotted lines indicate the locations of the densities' means. The mean of OD is higher than the mean of ARND by 0.71% (annualized 8.57%). The bottom panel plots option prices corresponding to ARND (the solid lines) and OD (the dashed lines), as functions of k. The upward-sloping curves are puts $p_t(k)$, the downward-sloping curves are calls $c_t(k)$. The dotted lines are the no-arbitrage bounds.

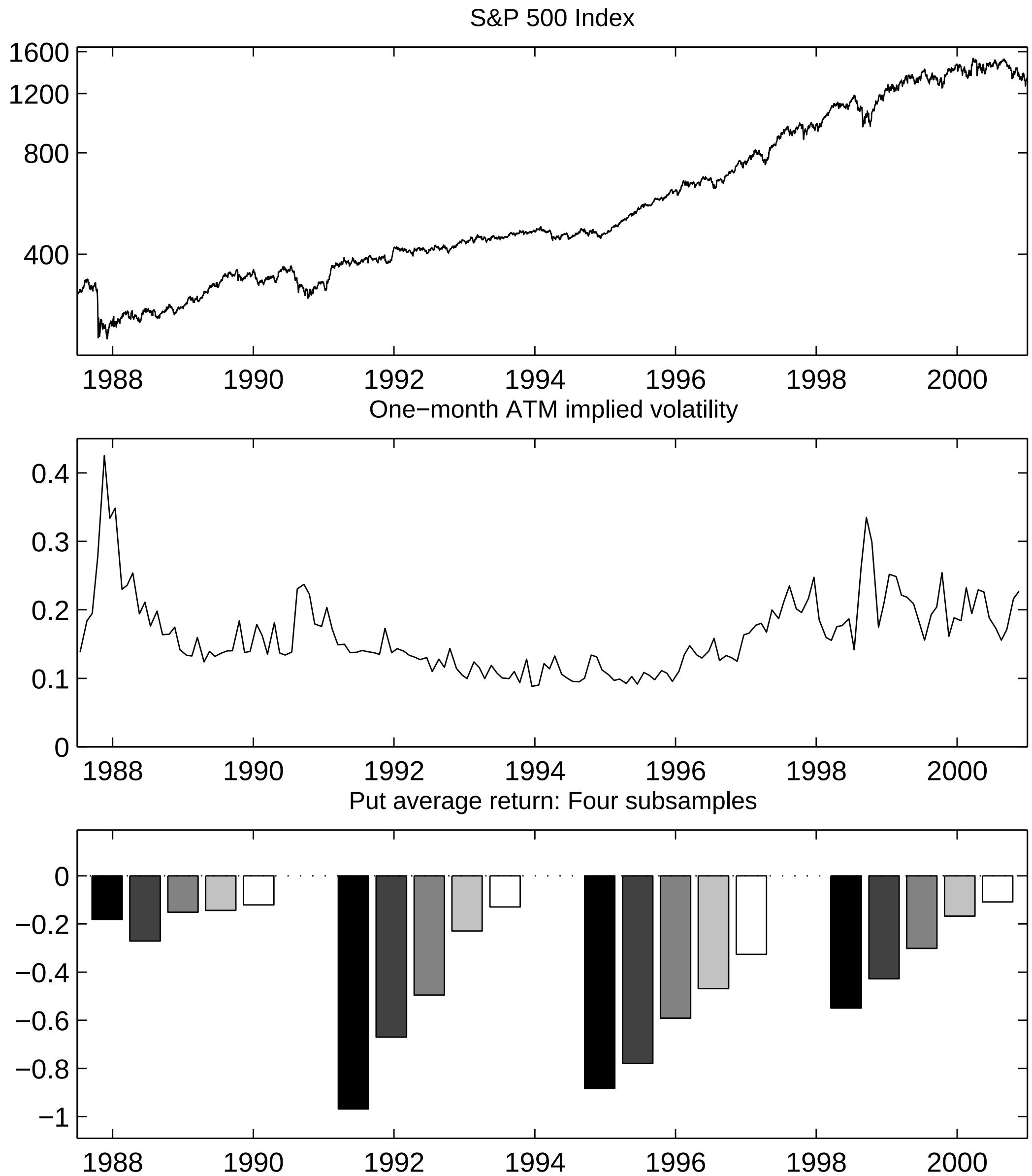

Figure 3: The top two panels plot the level of the S&P 500 Index and one-month ATM implied volatility from 08/87 to 12/00. The ATM volatility is annualized. The bottom panel shows put average returns over four subperiods: 08/87–06/90, 07/90–12/93, 01/94–06/97, and 07/97–12/00. The monthly average returns are shown for $k = 0.96$ (black bars), 0.98, 1.00, 1.02, and 1.04 (white bars).

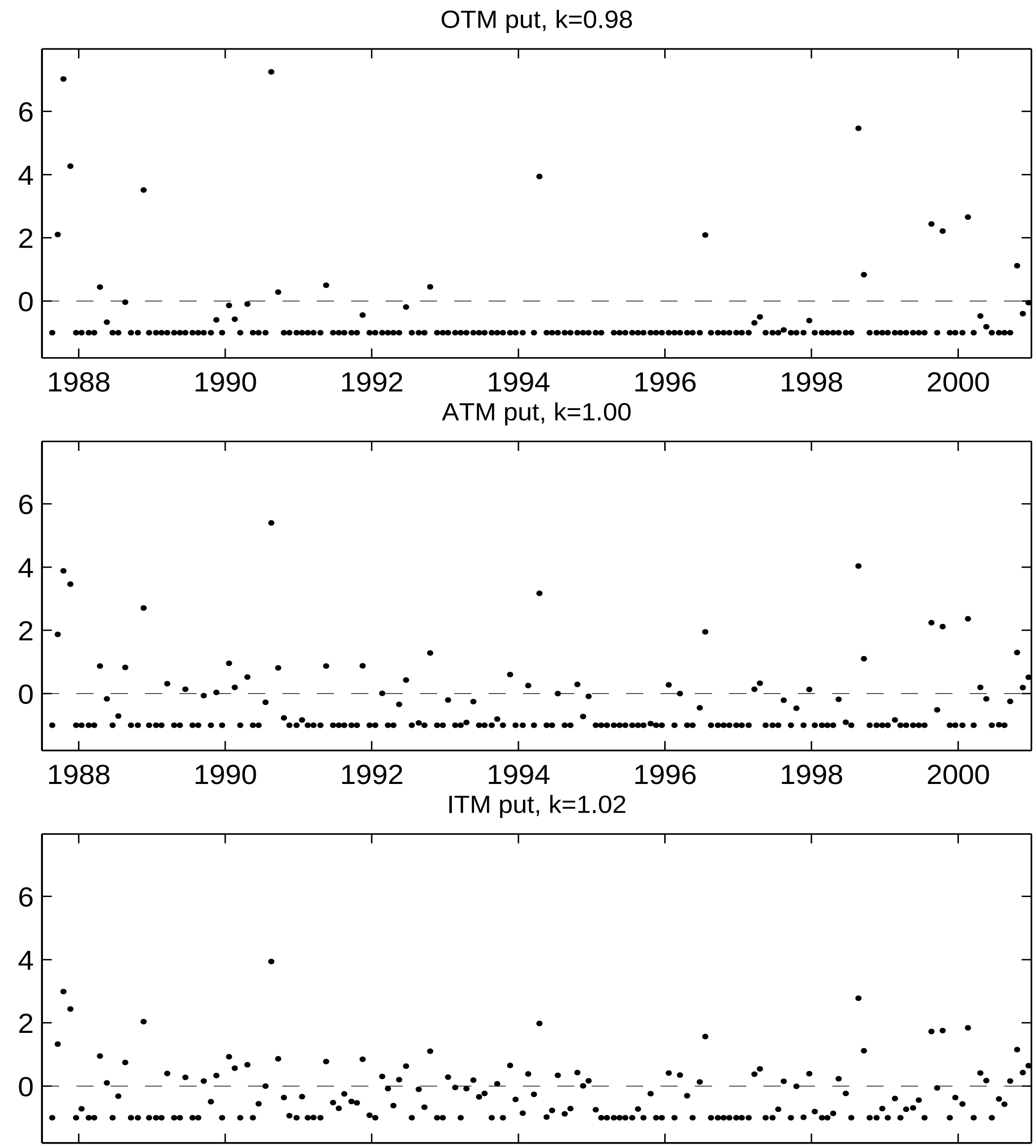

Figure 4: Put returns over time. Shown are OTM, ATM, and ITM options with $k = 0.98$, 1.00, and 1.02.

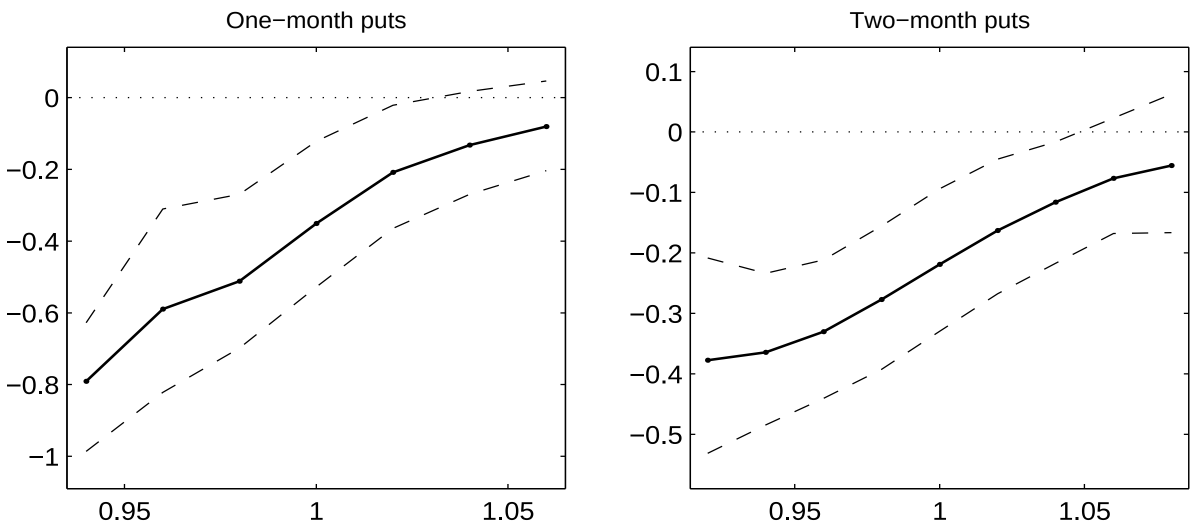

Figure 5: Monthly average weighted return AWR for one-month put (the left panel) and two-month puts (the right panel) with different moneyness k. Options are purchased 2 months and sold 1 month to prior the maturity date. The dashed lines are 5% and 95% pointwise confidence intervals, constructed using the bootstrap with 1000 resamples.

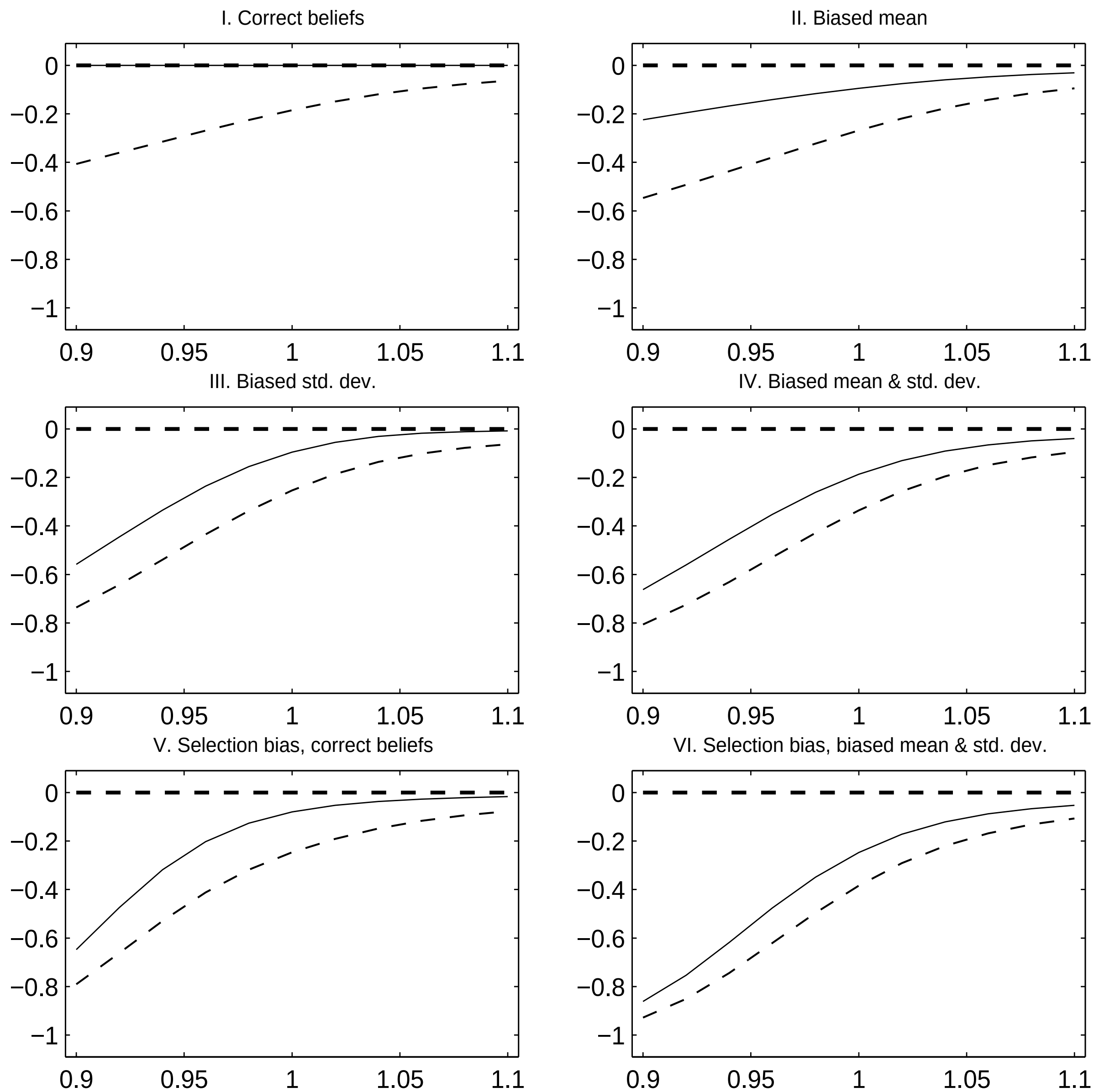

Figure 6: This figure illustrates the parametric example in Appendix C. The illustration assumes that $\sigma_0 = \sigma\sqrt{T}$, $\phi_t = \sigma(T-t)$, $\sigma = 0.04$, $\gamma = 4$, $s = 1$, $T = 2$, and $u_0 = \nu_0 + (\Delta + \gamma)\sigma_0^2$. The first four panels show the moments $I_1 = E_t[r_p(k)]$ (the thin dashed line), $I_2 = E_t[mr_p(k)]$ (the solid line), and $I_3^x = E_t^x[\lambda^x r_p(k)]$ (the thick dashed line) as functions of moneyness k for four special cases: Correct beliefs, $\Delta = 0$, $\eta_0 = \sigma_0$; Incorrect beliefs with a biased mean $\Delta = 2$, $\eta_0 = \sigma_0$; Incorrect beliefs with a biased standard deviation $\Delta = 0$, $\eta_0 = 0.85\sigma_0$; Incorrect beliefs with biased mean and standard deviation $\Delta = 2$, $\eta_0 = 0.85\sigma_0$. The last two panels show the moments $\bar{I}_1$, (the thin dashed line), $\bar{I}_2$, (the solid line), and $\bar{I}_3$ (the thick dashed line) as functions of moneyness k when 2% of the asset's lowest returns are discarded, for correct and incorrect beliefs.