

Understanding Index Option Returns

Mark Broadie, Mikhail Chernov, and Michael Johannes*

First Draft: September 2006

This Revision: January 8, 2008

Abstract

Previous research concludes that options are mispriced based on the high average returns, CAPM alphas, and Sharpe ratios of various put selling strategies. One criticism of these conclusions is that these benchmarks are ill-suited to handle the extreme statistical nature of option returns generated by nonlinear payoffs. We propose an alternative way to evaluate the statistical significance of option returns by comparing historical statistics to those generated by well-accepted option pricing models. The most puzzling finding in the existing literature, the large returns to writing out-of-the-money puts, are not inconsistent (i.e., are statistically insignificant) relative to the Black-Scholes model or the Heston stochastic volatility model due to the extreme sampling uncertainty associated with put returns. This sampling problem can largely be alleviated by analyzing market-neutral portfolios such as straddles or delta-hedged returns. The returns on these portfolios can be explained by jump risk premia and estimation risk.

*Broadie and Johannes are at Graduate School of Business, Columbia University. Chernov is at London Business School and CEPR. We thank David Bates, Alessandro Beber, Oleg Bondarenko, Peter Bossaerts, Pierre Collin-Dufresne, Kent Daniel, Joost Driessen, Bernard Dumas, Silverio Foresi, Vito Gala, Toby Moskowitz, Lasse Pedersen, Mirela Predescu, Todd Pulvino, Alex Reyfman, Alessandro Sbuelz and Christian Schlag for helpful comments. This paper was presented at the Adam Smith Asset Pricing conference, the 2008 AFA meetings in New Orleans, the Amsterdam Business School, AQR Capital Management, College of Queen Mary, Columbia, the ESSFM meetings in Gerzensee, Goldman Sachs Asset Management, HEC-Lausanne, HEC-Montreal, Lugano, Manchester Business School, Minnesota, the NBER Summer Institute, Tilburg, Universidade Nova de Lisboa, and the Yale School of Management. We thank Sam Cheung, Sudarshan Gururaj, and Pranay Jain for research assistance. Broadie acknowledges support by NSF grant DMS-0410234. Chernov acknowledges support of the BNP Paribas Hedge Funds Centre. Broadie and Chernov acknowledge support of the JP Morgan Chase Academic Outreach program.

1 Introduction

It is a common perception that index options are mispriced, in the sense that certain option returns are excessive relative to their risks.¹ The primary evidence supporting mispricing is the large magnitude of historical S&P 500 put option returns. For example, Bondarenko (2003) reports that average at-the-money (ATM) put returns are -40% , not per annum, but per month, and deep out-of-the-money (OTM) put returns are -95% per month. Average option returns and CAPM alphas are statistically significant with p -values close to zero, and Sharpe ratios are larger than those of the underlying index.²

There are three obvious factors to consider when interpreting these results. Option returns are highly non-normal and metrics that assume normality, such as CAPM alphas or Sharpe ratios, are inappropriate. In addition, average put returns or CAPM alphas should be significantly different from zero due to the leverage inherent in options and the presence of priced risks that primarily affect higher moments such as jumps. Finally, options have only traded for a short period of time, and it is difficult to assess the statistical significance of option returns given these short time spans and the non-normal nature of option returns. Together, these factors raise questions about the usual procedures of applying standard asset pricing metrics to analyze option returns.

A natural way to deal with these criticisms is to use option pricing models to assess the evidence for index option mispricing. Option models automatically account for the extreme nature of option returns (non-normality, skewness and fat-tails), anchor hypothesis tests at appropriate null values, provide a framework for assessing the impact of risk premia, and provide a mechanism for assessing statistical uncertainty via finite sample simulations. Ideally, an equilibrium model over economic fundamentals, such as consumption or dividends, would be used to assess the evidence for mispricing. However, as argued by Bates (2006) and Bondarenko (2003), such an explanation is extremely challenging inside the

¹At this stage, a natural question to ask is why returns and why not option prices? Throughout finance, returns, as opposed to price levels, are typically analyzed because of their natural economic interpretation. Returns represent actual gains or losses on purchased securities. In contrast, common option pricing exercises use pricing errors to summarize fit, which are neither easily interpreted nor can be realized. In addition, we have stronger intuition about return-based measures such as excess returns, CAPM alphas, or Sharpe ratios as compared to pricing errors. Coval and Shumway (2000) provide additional motivation.

²The returns are economically significant, as investors endowed with a wide array of utility functions find large certainty equivalent gains from selling put options (e.g., Driessen and Maenhout, 2004; Santa-Clara and Saretto, 2005).

representative agent framework. This conclusion is not surprising, since these models have difficulties explaining not only the low-frequency features of stock returns (e.g., the equity premium or excess volatility puzzles), but also higher frequency movements such as price jumps, high-frequency volatility fluctuations, or the leverage effect. At some level, these equilibrium models do not operate at a frequency that is relevant for option pricing.

This paper addresses a more modest, but still important, goal of understanding the pricing of index options relative to the underlying index, as opposed to pricing options relative to the underlying fundamental variables. To do this, we model stock index returns using affine-jump diffusion models that account for the key drivers of equity returns and option prices such as diffusive price shocks, price jumps, and stochastic volatility. The key step in our implementation is one of calibration: we calibrate the models to fit the observed behavior of equity index returns over the sample for which option returns are available. In particular, this approach implies that our models replicate the historically observed equity premium and volatility.

Methodologically, we proceed using two main tools. First, we show that expected option returns (EORs) can be computed analytically, which allows us to examine the quantitative implications of different factors and parameter values on option returns. In particular, EORs anchor null hypothesis values when testing whether option returns are significantly different than those generated by a given null model. Second, simulated index sample paths are used to construct exact finite sample distributions for the statistics analyzed. This procedure accounts for the small observed samples sizes (on the order of 200 months) and the irregular nature of option return distributions. Another advantage is that it allows us to assess the statistical uncertainty of commonly used asset pricing benchmarks and statistics, such as average returns, CAPM alphas, or Sharpe ratios, while accounting for the leverage and nonlinearities inherent in options.

Empirically, we present a number of interesting findings. We first analyze returns on individual put options, given their importance in the recent literature, and begin with the simplest Black-Scholes and Heston (1993) stochastic volatility models. Although we know that these models are too simple to provide accurate descriptions of option *prices*, they provide key insights for understanding and evaluating option *returns*. Our first result is that average returns, CAPM alphas, and Sharpe ratios for deep OTM put returns, are statistically *insignificant* when compared to the Black-Scholes model. Thus, one of the most puzzling statistics in the literature, the high average returns on OTM puts, is not

inconsistent with the Black-Scholes model. Moreover, there is little evidence that put returns of any strike are inconsistent with Heston's (1993) stochastic volatility (SV) model assuming no diffusive stochastic volatility risk premia (i.e., the evolution of volatility under the real-world \mathbb{P} and the risk-neutral \mathbb{Q} measures are the same).

We interpret these findings not as evidence that Black-Scholes or Heston's models are correct – we know they can be rejected as models of option prices on other grounds – but rather as highlighting the statistical difficulties present when analyzing option returns. The combination of short samples and complicated option return distributions implies that standard statistics are so noisy that little can be concluded by analyzing option returns. It is well known that it is difficult to estimate the equity premium, and this uncertainty is magnified when estimating average put returns.

This conclusion suggests that tests using individual put option returns are not very informative about option mispricing, and we next turn to returns of alternative option portfolio strategies such as covered puts, delta-hedged puts, put spreads, at-the-money straddles, and crash-neutral straddles. These portfolios are more informative because they either reduce the exposure to the underlying index (delta-hedged puts and straddles) or dampen the effect of rare events (put spreads and crash-neutral straddles).

Of all of these option portfolios, straddles returns are the most useful as they are model independent and approximately market neutral. Straddles are also more informative than individual puts, since average straddle returns are highly significant when compared to returns generated from the Black-Scholes model or a baseline stochastic volatility model without a diffusive volatility risk premium. The source of the significance for ATM straddle returns is the well-known wedge between ATM implied volatility and subsequent realized volatility.³ As argued by Pan (2002) and Broadie, Chernov, and Johannes (2007), it is unlikely that a diffusive stochastic volatility risk premium could generate this wedge between \mathbb{Q} and \mathbb{P} measures since the wedge is in short-dated options, and a stochastic volatility risk premium would mainly impact longer-dated options.

We consider two mechanisms that generate wedges between realized and implied volatility in jump-diffusion models: jump risk premia and estimation risk. The jump risk explanation uses the jump risk corrections implied by an equilibrium model as a simple device for generating \mathbb{Q} -measure jump parameters, given \mathbb{P} -measure parameters. Consistent with our original intent, such an adjustment does not provide an equilibrium explanation, as we

³Over our sample, ATM implied volatility averaged 17% and realized volatility was 15%.

calibrate the underlying index model to match the overall equity premium and volatility of returns. In the case of estimation risk, we assume that investors account for the uncertainty in spot volatility and parameters when pricing options.

We find that both of these explanations generate option returns that are broadly consistent with those observed historically. For example, average put returns are matched pointwise and the average returns of straddles, delta-hedged portfolios, put spreads, and crash-neutral straddles are all statistically insignificant. These results indicate that, at least for our parameterizations, that option returns are not puzzling relative to the benchmark models. Option and stock returns may remain puzzling relative to consumption and dividends, but there appears to be little evidence for mispricing relative to the underlying stock index.

The rest of the paper is outlined as follows. Section 2 outlines our methodological approach. Section 3 discusses our data set and summarizes the evidence for put mispricing. Section 4 illustrates the methodology based on the Black-Scholes and Heston models. Section 5 investigates strategies based on option portfolios. Section 6 illustrates how a model with stochastic volatility and jumps in prices generate realistic put and straddle returns. Conclusions are given in Section 7.

2 Our approach

We analyze returns to a number of option strategies. In this section, we discuss some of the concerns that arise in analyzing option returns, and then discuss our approach. To frame the issues, we focus on put option returns, but the results and discussion apply more generally to portfolio strategies such as put spreads, straddles, or delta-hedged returns.

Hold-to-expiration put returns are defined as

$$r_{t,T}^p = \frac{(K - S_{t+T})^+}{P_{t,T}(K, S_t)} - 1, \quad (2.1)$$

where $x^+ \equiv \max(x, 0)$ and $P_{t,T}(K, S_t)$ is the time- t price of a put option written on S_t , struck at K , and expiring at time $t + T$. Hold-to-expiration returns are typically analyzed in both academic studies and in practice for two reasons. First, option trading involves significant costs and strategies that hold until expiration incur these costs only at initiation. For example, ATM (deep OTM) index option bid-ask spreads are currently on the order of

3%–5% (10%) of the option price. The second reason, discussed fully in Section 3, is that higher frequency option returns generate a number of theoretical and statistical issues that are avoided using monthly returns.

The main objective in the literature is assessing whether or not option returns are excessive, either in absolute terms or relative to their risks. Existing approaches rely on statistical models, as discussed in Appendix A. For example, it is common to compute average returns, alphas relative to the CAPM, or Sharpe ratios. Strategies that involve writing options generally deliver higher average returns than the underlying asset, have economically and statistically large CAPM alphas, and have higher Sharpe ratios than the market.

How should these results be interpreted? Options are effectively leveraged positions in the underlying asset (which typically has a positive expected return), so call options have expected returns that are greater than the underlying and put options have expected returns that are less than the underlying. For example, expected put option returns are negative, which implies that standard t -tests of average option returns which test the null hypothesis that average returns are zero are not particularly informative. The precise magnitude of expected returns depends on a number of factors that include the specific model, the parameters, and factor risk premia. In particular, expected option returns are very sensitive to both the equity premium and volatility.

It is important to control for the option's exposure to the underlying, and the most common way to do this is to compute betas relative to the underlying asset via a CAPM-style specification. This approach is motivated by the hedging arguments used to derive the Black-Scholes model. According to this model, the link between instantaneous derivative returns and excess index returns is

$$\frac{df(S_t)}{f(S_t)} = rdt + \frac{S_t}{f(S_t)} \frac{\partial f(S_t)}{\partial S_t} \left[\frac{dS_t}{S_t} - (r - \delta) dt \right], \quad (2.2)$$

where r is the risk-free rate, $f(S_t)$ is the derivative price, and δ the dividend rate on the underlying asset. This implies that instantaneous changes in the derivative's price are *linear* in the index returns, dS_t/S_t , and *instantaneous* option returns are conditionally normally distributed. This instantaneous CAPM is often used to motivate an *approximate* linear factor model for option returns

$$\frac{f(S_{t+T}) - f(S_t)}{f(S_t)} = \alpha_{t,T} + \beta_{t,T} \left(\frac{S_{t+T} - S_t}{S_t} - rT \right) + \varepsilon_{t,T}.$$

These linear factor models are used to adjust for leverage, via $\beta_{t,T}$, and as a pricing metric, via $\alpha_{t,T}$. In the latter case, $\alpha_T \neq 0$ is often interpreted as evidence of either mispricing or risk premia.

As shown in detail in Appendix B, standard option pricing models (including the Black-Scholes model) generate population values of $\alpha_{t,T}$ that are different from zero. In the Black-Scholes model, this is due to time discretization, but in more general jump-diffusion models, $\alpha_{t,T}$ can be non-zero for infinitesimal intervals due to the presence of jumps. This implies that it is inappropriate to interpret a non-zero $\alpha_{t,T}$ as evidence of mispricing. Similarly, Sharpe ratios account for leverage by scaling average excess returns by volatility, which provides an appropriate metric when returns are normally distributed or if investors have mean-variance preferences. Sharpe ratios are problematic in our setting because option returns are highly non-normal, even over short time-intervals.

Our approach is different. We view these intuitive metrics (average returns, CAPM alpha's, and Sharpe ratios) through the lens of formal option pricing models. The experiment we perform is straightforward: we compare the observed values of these statistics in the data to those generated by option pricing models such as Black-Scholes and extensions incorporating jumps or stochastic volatility. The use of formal models performs two roles: it provides an appropriate null value for anchoring hypothesis tests and it provides a mechanism for dealing with the severe statistical problems associated with option returns.

2.1 Models

We consider nested versions of a general model with mean-reverting stochastic volatility and lognormally distributed Poisson driven jumps in prices. This model, proposed by Bates (1996) and Scott (1997) and referred to as the SVJ model, is a common benchmark (see, e.g., Andersen, Benzoni, and Lund (2002), Bates (1996), Broadie, Chernov, and Johannes (2007), Chernov, Gallant, Ghysels, and Tauchen (2003), Eraker (2004), Eraker, Johannes, and Polson (2003), and Pan (2002)). As special cases of the model, we consider the Black and Scholes (1973) model, Merton's (1976) jump-diffusion model with constant volatility, and Heston's (1993) stochastic volatility model.

The model assumes that the ex-dividend index level, S_t , and its spot variance, V_t , evolve

under the physical (or real-world) \mathbb{P} -measure according to

$$dS_t = (r + \mu - \delta) S_t dt + S_t \sqrt{V_t} dW_t^s(\mathbb{P}) + d \left(\sum_{j=1}^{N_t(\mathbb{P})} S_{\tau_{j-}} \left[e^{Z_j^s(\mathbb{P})} - 1 \right] \right) - \lambda^{\mathbb{P}} \bar{\mu}^{\mathbb{P}} S_t dt \quad (2.3)$$

$$dV_t = \kappa_v^{\mathbb{P}} (\theta_v^{\mathbb{P}} - V_t) dt + \sigma_v \sqrt{V_t} dW_t^v(\mathbb{P}), \quad (2.4)$$

where r is the risk-free rate, μ is the cum-dividend equity premium, δ is the dividend yield, W_t^s and W_t^v are two correlated Brownian motions ($E[W_t^s W_t^v] = \rho t$), $N_t(\mathbb{P}) \sim \text{Poisson}(\lambda^{\mathbb{P}} t)$, $Z_j^s(\mathbb{P}) \sim \mathcal{N}(\mu_z^{\mathbb{P}}, (\sigma_z^{\mathbb{P}})^2)$, and $\bar{\mu}^{\mathbb{P}} = \exp(\mu_z^{\mathbb{P}} + (\sigma_z^{\mathbb{P}})^2/2) - 1$. Black-Scholes is a special case with no jumps ($\lambda^{\mathbb{P}} = 0$) and constant volatility ($V_0 = \theta_v^{\mathbb{P}}$, $\sigma_v = 0$), Heston's model is a special case with no jumps, and Merton's model is a special case with constant volatility. When volatility is constant, we use the notation $\sqrt{V_t} = \sigma$.

Options are priced using the dynamics under the risk-neutral measure \mathbb{Q} :

$$dS_t = (r - \delta) S_t dt + S_t \sqrt{V_t} dW_t^s(\mathbb{Q}) + d \left(\sum_{j=1}^{N_t(\mathbb{Q})} S_{\tau_{j-}} \left[e^{Z_j^s(\mathbb{Q})} - 1 \right] \right) - \lambda^{\mathbb{Q}} \bar{\mu}^{\mathbb{Q}} S_t dt \quad (2.5)$$

$$dV_t = \kappa_v^{\mathbb{Q}} (\theta_v^{\mathbb{Q}} - V_t) dt + \sigma_v \sqrt{V_t} dW_t^v(\mathbb{Q}), \quad (2.6)$$

where $N_t(\mathbb{Q}) \sim \text{Poisson}(\lambda^{\mathbb{Q}} t)$, $Z_j(\mathbb{Q}) \sim \mathcal{N}(\mu_z^{\mathbb{Q}}, (\sigma_z^{\mathbb{Q}})^2)$, $W_t(\mathbb{Q})$ are Brownian motions, and $\bar{\mu}^{\mathbb{Q}}$ is defined analogously to $\bar{\mu}^{\mathbb{P}}$. The diffusive equity premium is μ^c , and the total equity premium is $\mu = \mu^c + \lambda^{\mathbb{P}} \bar{\mu}^{\mathbb{P}} - \lambda^{\mathbb{Q}} \bar{\mu}^{\mathbb{Q}}$. We generally refer to a non-zero μ as a diffusive risk premium. Differences between the risk-neutral and real-world jump and stochastic volatility parameters are referred to as jump or stochastic volatility risk premia, respectively.

The parameters θ_v and κ_v can both potentially change under the risk-neutral measure (Cheredito, Filipovic, and Kimmel (2003)). We explore changes in $\theta_v^{\mathbb{P}}$ and constrain $\kappa_v^{\mathbb{Q}} = \kappa_v^{\mathbb{P}}$, because, as discussed below, average returns are not sensitive to empirically plausible changes in $\kappa_v^{\mathbb{P}}$. Changes of measure for jump processes are more flexible than those for diffusion processes. We take the simplifying assumptions that the jump size distribution is lognormal with potentially different means and variances. Below we discuss in detail two mechanisms, risk premia and estimation risk, to generate realistic \mathbb{Q} -measure parameters.

2.2 Methodological framework

Methodologically, we rely on two main tools: analytical formulas for expected returns and Monte Carlo simulation to assess statistical significance.

2.2.1 Analytical expected option returns

Expected put option returns are given by

$$\begin{aligned} E_t^{\mathbb{P}}(r_{t,T}^p) &= E_t^{\mathbb{P}} \left[\frac{(K - S_{t+T})^+}{P_{t,T}(S_t, K)} \right] - 1 = \frac{E_t^{\mathbb{P}} [(K - S_{t+T})^+]}{P_{t,T}(S_t, K)} - 1 \\ &= \frac{E_t^{\mathbb{P}} [(K - S_{t+T})^+]}{E_t^{\mathbb{Q}} [e^{-rT} (K - S_{t+T})^+]} - 1, \end{aligned} \quad (2.7)$$

where the second equality emphasizes that $P_{t,T}$ is known at time t . Put prices will depend on the specific model under consideration. From this expression, it is clear that any model that admits “analytical” option prices, such as affine models, will allow EORs to be computed explicitly since both the numerator and denominator are known analytically. Higher moments can also be computed. Surprisingly, despite a large literature analyzing option returns, the fact that EORs can be easily computed has neither been noted nor applied.⁴

EORs do not depend on S_t . To see this, define the initial moneyness of the option as $\kappa = K/S_t$. Option homogeneity implies that

$$E_t^{\mathbb{P}}(r_{t,T}^p) = \frac{E_t^{\mathbb{P}} [(\kappa - R_{t,T})^+]}{E_t^{\mathbb{Q}} [e^{-rT} (\kappa - R_{t,T})^+]} - 1, \quad (2.8)$$

where $R_{t,T} = S_{t+T}/S_t$ is the gross index return. Expected option returns depend only on the moneyness, maturity, interest rate, and the distribution of index returns.⁵

This formula provides exact EORs for finite holding periods regardless of the risk factors of the underlying index dynamics, without using CAPM-style approximations such as those discussed in Appendix B. These analytical results are primarily useful as they allow us to assess the exact *quantitative* impact of risk premia or parameter configurations. Equation (2.7) implies that the gap between the \mathbb{P} and \mathbb{Q} probability measures determines expected option returns, and the magnitude of the returns is determined by the relative

⁴This result is closely related to Rubinstein (1984), who derived it specifically for the Black-Scholes case and analyzed the relationship between hold-to-expiration and shorter holding period expected returns.

⁵When stochastic volatility is present in a model, the expected option returns can be computed analytically conditional on the current variance value: $E^{\mathbb{P}}(r_{t,T}^p|V_t)$. The unconditional expected returns can be computed using iterated expectations and the fact that

$$E^{\mathbb{P}}(r_{t,T}^p) = \int E^{\mathbb{P}}(r_{t,T}^p|V_t) p(V_t) dV_t.$$

The integral can be estimated via Monte Carlo simulation or by standard deterministic integration routines.

shape and location of the two probability measures.⁶ In models without jump or stochastic volatility risk premia, the gap is determined by the fact that the \mathbb{P} and \mathbb{Q} drifts differ by the equity risk premium. In models with priced stochastic volatility or jump risk, both the shape and location of the distribution can change, leading to more interesting patterns of expected returns across different moneyness categories.

2.2.2 Finite sample distribution via Monte Carlo simulation

To assess statistical significance, we use Monte Carlo simulation to compute the distribution of various returns statistics, including average returns, CAPM alphas, and Sharpe ratios. We are motivated by concerns that the use of limiting distributions to approximate the finite sample distribution is inaccurate in this setting. Our concerns arise due to the relatively short sample and extreme skewness and non-normality of option returns.

To compute the finite sample distribution of various option return statistics, we simulate N months (the sample length in the data) of index levels $G = 25,000$ times using standard simulation techniques. For each month and index simulation trial, put returns are

$$r_{t,T}^{p,(g)} = \frac{(\kappa - R_{t,T}^{(g)})^+}{P_T(\kappa)} - 1, \quad (2.9)$$

where

$$P_T(\kappa) \triangleq \frac{P_{t,T}(S_t, K)}{S_t} = e^{-rT} E_t^{\mathbb{Q}} [(\kappa - R_{t,T})^+],$$

$t = 1, \dots, N$ and $g = 1, \dots, G$. Average option returns over the N months on simulation trial g are given by

$$\bar{r}_T^{p,(g)} = \frac{1}{N} \sum_{t=1}^N r_{t,T}^{p,(g)}.$$

A set of G average returns forms the finite sample distribution. Similarly, we can construct finite sample distributions for Sharpe ratios, CAPM alphas, and other statistics of interest for any option portfolio.

This parametric bootstrapping approach provides exact finite sample inference under the null hypothesis that a given model holds. It can be contrasted with the nonparametric bootstrap, which creates artificial datasets by sampling with replacement from the observed data. The nonparametric bootstrap, which essentially reshuffles existing observations, has

⁶For monthly holding periods, $1 \leq \exp(rT) \leq 1.008$ for $0\% \leq r \leq 10\%$ and $T = 1/12$ years, so the discount factor has a negligible impact on EORs.

difficulties dealing with rare events. In fact, if an event has not occurred in the observed sample, it will never appear in the simulated finite sample distribution. This is an important concern when dealing with put returns which are very sensitive to rare events.

2.3 Parameter estimation

We calibrate our models to fit the realized historical behavior of the underlying index returns over our observed sample. Thus, the \mathbb{P} -measure parameters are estimated directly from historical index return data, and not consumption or dividend behavior. For parameters in the Black-Scholes model, this calibration is straightforward, but in models with unobserved volatility or jumps, the estimation is more complicated as it is not possible to estimate all of the parameter via simple sample statistics.

For all of the models that we consider, the interest rate and equity premium match those observed over our sample, $r = 4.5\%$ and $\mu = 5.4\%$. Since we analyze futures returns and futures options, $\delta = r$. In each model, we also constrain the total volatility to match the observed monthly volatility of futures returns, which was 15%. In the most general model we consider, we do this by imposing that

$$\sqrt{\theta_v^{\mathbb{P}} + \lambda^{\mathbb{P}} ((\mu_z^{\mathbb{P}})^2 + (\sigma_z^{\mathbb{P}})^2)} = 15\%$$

and by modifying $\theta_v^{\mathbb{P}}$ appropriately. In the Black-Scholes model, we set the constant volatility to be 15%.

To obtain the values of the remaining parameters, we estimate the SVJ model using daily S&P 500 index returns spanning the same time period as our options data, from 1987 to 2005. We use MCMC methods to simulate the posterior distribution of the parameters and state variables following Eraker, Johannes, and Polson (2003) and others. The parameter estimates (posterior means) and posterior standard deviations are reported in Table 1. The parameter estimates are in line with the values reported in previous studies (see Broadie, Chernov, and Johannes, 2007 for a review).

Of particular interest for our analysis are the jump parameters. The estimates imply that jumps are relatively infrequent, arriving at a rate of about $\lambda^{\mathbb{P}} = 0.91$ per year. The jumps are modestly sized with the mean of -3.25% and a standard deviation of 6%. Given these values, a “two sigma” downward jump size will be equal to -15.25% . Therefore, a crash-type move of -15% , or below, will occur with a probability of $\lambda^{\mathbb{P}} \cdot 5\%$, or, approximately once in twenty years.

r	μ	$\lambda^{\mathbb{P}}$	$\mu_z^{\mathbb{P}}$	$\sigma_z^{\mathbb{P}}$	$\sqrt{\theta_v^{\mathbb{P}}}$ (SV)	$\sqrt{\theta_v^{\mathbb{P}}}$ (SVJ)	$\kappa_v^{\mathbb{P}}$	σ_v	ρ
4.50%	5.41%	0.91	-3.25%	6.00%	15.00%	13.51%	5.33	0.14	-0.52
		(0.34)	(1.71)	(0.99)		(1.28)	(0.84)	(0.01)	(0.04)

Table 1: \mathbb{P} -measure parameters. We report parameter values that we use in our computational examples. Standard errors from the SVJ estimation are reported in parentheses. Parameters are given in annual terms.

As we discuss in greater detail below, estimating jump intensities and jump size distributions is extremely difficult. The estimates are highly dependent on the observed data and on the specific model. For example, different estimates would likely be obtained if we assumed that the jump intensity was dependent on volatility (as in Bates (2000) or Pan (2002)) or if there were jumps in volatility. Again, our goal is not to exhaustively analyze every potential specification, but rather to understand option returns in common specifications and for plausible parameter values.

We discuss the calibration of \mathbb{Q} -measure parameters later. At this stage, we only emphasize that we do not use options data to estimate any of the parameters. Estimating \mathbb{Q} -parameters from option prices for use in understanding observed option returns would introduce a circularity, as we would be explaining option returns with information extracted from option prices.

3 Initial evidence for put mispricing

We collect historical data on S&P 500 futures options from August 1987 to June 2005, a total of 215 months. This sample is considerably longer than those previously analyzed and starts in August of 1987 when one-month “serial” options were introduced. Contracts expire on the third Friday of each month, which implies there are 28 or 35 calendar days to maturity depending on whether it was a four- or five-week month. We construct representative daily option prices using the approach in Broadie, Chernov, and Johannes (2007); details of this procedure are given in Appendix C.

Using these prices, we compute option returns for fixed moneyness, measured by strike

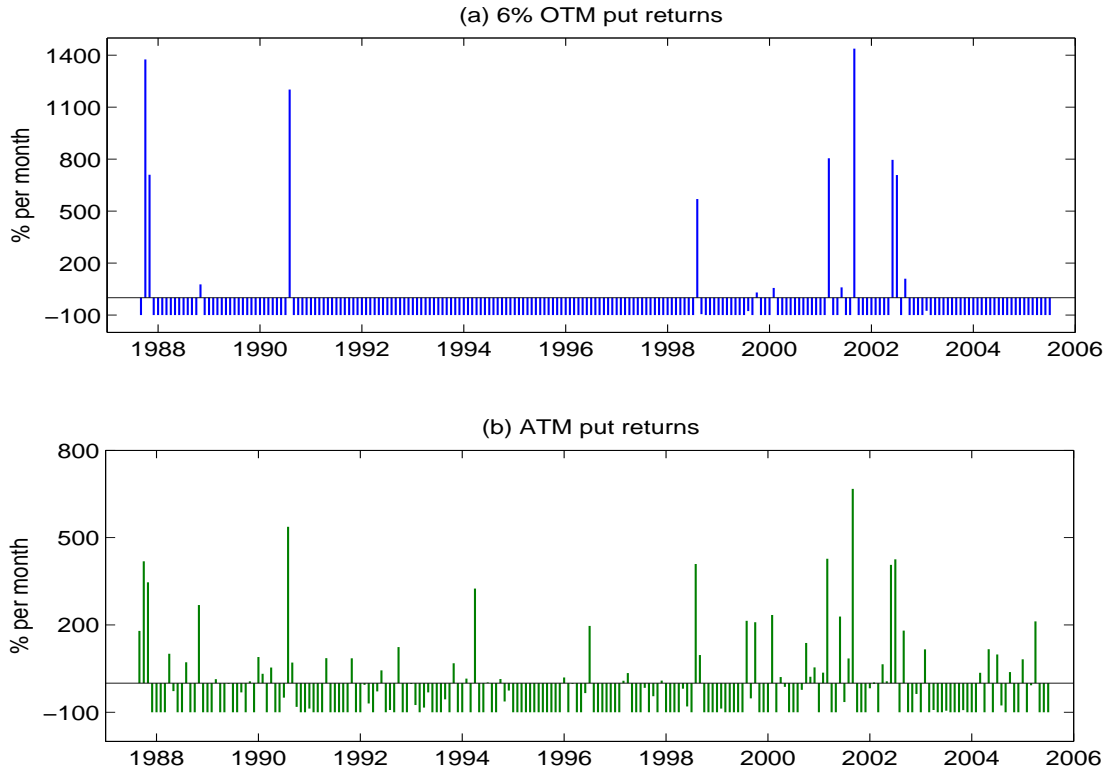


Figure 1: Time series of options returns.

divided by the underlying, ranging from 0.94 to 1.02 (in 2% increments), which represents the most actively traded options (85% of one-month option transactions occur in this range). We did not include deeper OTM or ITM strikes because of missing values. Payoffs are computed using settlement values for the S&P 500 futures contract. Figure 1 shows the time series for 6% OTM and ATM put returns, which highlights some of the issues that are present when evaluating the statistical significance of statistics generated by option returns. The put return time series have very large outliers and many repeated values, since OTM expirations generate returns of -100% . We also compute returns for a range of portfolio strategies, including covered puts, put spreads (crash-neutral put portfolios), delta-hedged puts, straddles, and crash-neutral straddles. For clarity, we first consider the returns to writing put options, as this has been the primary focus in the existing literature.

Moneyness	0.94	0.96	0.98	1.00	1.02
08/1987 to 06/2005	-56.8	-52.3	-44.7	-29.9	-19.0
Standard error	14.2	12.3	10.6	8.8	7.1
t -stat	-3.9	-4.2	-4.2	-3.3	-2.6
p -value, %	0.0	0.0	0.0	0.0	0.4
Skew	5.5	4.5	3.6	2.5	1.8
Kurt	34.2	25.1	16.7	10.5	7.1
Subsamples					
01/1988 to 06/2005	-65.2	-60.6	-51.5	-34.1	-21.6
01/1995 to 09/2000	-85.5	-71.6	-63.5	-50.5	-37.5
10/2000 to 02/2003	+67.2	+54.3	+44.5	+48.2	+40.4
08/1987 to 01/2000	-83.9	-63.2	-55.7	-39.5	-25.5

Table 2: Average put option returns. The first panel contains the full sample, with standard errors, t -statistics, and skewness and kurtosis statistics. The second panel analyzes subsamples. All relevant statistics are in percentages per month.

As mentioned earlier, we focus on hold-to-maturity returns. The alternative would be higher frequency returns, such as weekly or even daily. The intuition for considering higher frequency returns comes from the Black-Scholes dynamic hedging arguments indicating that option returns become approximately normal over high frequencies. Appendix D describes the difficulties associated with higher frequency returns in detail. In particular, we argue that using higher frequency returns generates additional theoretical, data, and statistical problems. In particular, simulation evidence shows that moving from monthly to weekly returns hurts rather than helps the statistical issues because the distribution of average returns becomes even more non-normal and dispersed.

Table 2 reports average put returns, standard errors, t -statistics, p -values, and measures of non-normality (skewness and kurtosis). We also report average returns over various subsamples.

The first piece of evidence commonly cited supporting mispricing is the large magnitude of the returns: average monthly returns are -57% for 6% OTM strikes (i.e., $K/S = 0.94$) and -30% for ATM strikes and are statistically different from zero using t -statistics, as p -values are close to zero. The bottom panel reports average returns over subsamples. In

particular, to check that our results are consistent with previous findings, we compare our statistics to the ones in the Bondarenko (2003) sample from 1987 to 2000. The returns are very close, but ours are slightly more negative for every moneyness category except the deepest OTM category. Bondarenko (2003) uses closing prices and has some missing values. Our returns are more negative than those reported for similar time periods by Santa-Clara and Saretto (2005).

Average put returns are unstable over time. For example, put returns were extremely negative during the late 1990s during the dot-com “bubble,” but were positive and large from late 2000 to early 2003. The subsample starting in January 1988 provides the same insight: if the extremely large positive returns realized around the crash of 1987 are excluded, returns are much lower. Doing so generates a sample selection bias and clearly demonstrates a problem with tests using short sample periods.⁷

Table 3 reports CAPM alphas and Sharpe ratios, which delever and/or risk-correct option returns to account for the underlying exposure. CAPM alphas are highly statistically significant, with p -values near zero. The Sharpe ratios of put positions are larger than those on the underlying market. For example, the monthly Sharpe ratio for the market over our time period was about 0.1, and the put return Sharpe ratios are two to three times larger. Based largely on this evidence and additional robustness checks, the literature concludes that put returns are puzzling and options are likely mispriced. We briefly review the related literature in Appendix A.

4 The role of statistical uncertainty

This section highlights the difficulties in analyzing potential option mispricing based on returns of individual options. We rely on the simplest option pricing models, that is, the Black-Scholes and SV models without a stochastic volatility risk premium. We show that expected returns are highly sensitive to the underlying equity premium and volatility and also document the extreme finite sample problems associated with tests using returns of individual options. In particular, the biggest puzzle in the literature – the large deep

⁷In simulations of the Black-Scholes model, excluding the largest positive return reduces *average* option returns by about 15% for the 6% OTM strike. This outcome illustrates the potential sample selection issues and how sensitive option returns are to the rare but extremely large positive returns generated by events such as the crash of 1987.

Moneyiness	0.94	0.96	0.98	1.00	1.02
CAPM α , %	-48.3	-44.1	-36.8	-22.5	-12.5
Std.err., %	11.6	9.3	7.1	4.8	2.9
t -stat	-4.1	-4.7	-5.1	-4.6	-4.2
p -value, %	0.0	0.0	0.0	0.0	0.0
Sharpe ratio	-0.27	-0.29	-0.29	-0.23	-0.18

Table 3: Risk-corrected measures of average put option returns. The first panel provides CAPM α 's with standard errors and the second panel provides put option Sharpe ratios. All relevant statistics except for the Sharpe ratios are in percentages per month. Sharpe ratios are monthly. The p -values are computed under the (incorrect) assumption that t -statistics are t -distributed.

OTM put returns – is not inconsistent with the Black-Scholes model because of statistical considerations.

4.1 Black-Scholes

In this section, we analyze expected option returns in the Black-Scholes model and analyze the finite sample distribution of average option returns, CAPM alphas, and Sharpe ratios. In the Black-Scholes model, EORs are large in magnitude, negative, and highly sensitive to the equity premium μ and volatility σ , especially for OTM strikes. To show this, Table 4 computes EORs using equation (2.8). The cum-dividend equity premium ranges from 4% to 8% and volatility ranges from 10% to 20%. The impact of μ is approximately linear and quantitatively large, as the difference in EORs between high and low equity premiums is about 10% for ATM strikes and more for deep OTM strikes. Because of this, any historical period that is “puzzling” because of high realized equity returns will generate option returns that are even more striking. For example, the realized equity premium from 1990 to 1999 was 9.4% and average volatility was only 13% over the same period. If fully anticipated, these values would, according to equation (2.8), generate 6% OTM and ATM EORs of about -40% and -23%, respectively, which are much lower than the EORs using the full sample equity premium and volatility.

Option returns are sensitive to volatility. As volatility increases, expected put option

		Moneyness						
σ	μ	0.94	0.96	0.98	1.00	1.02	1.04	1.06
10%	4%	-27.6	-22.5	-17.6	-13.3	-9.7	-6.9	-5.0
	6%	-38.7	-32.2	-25.7	-19.7	-14.5	-10.5	-7.7
	8%	-48.3	-40.8	-33.1	-25.7	-19.2	-14.1	-10.4
15%	4%	-15.4	-13.0	-10.8	-8.8	-7.1	-5.6	-4.5
	6%	-22.5	-19.2	-16.1	-13.2	-10.7	-8.6	-6.9
	8%	-29.1	-25.0	-21.1	-17.5	-14.3	-11.5	-9.3
20%	4%	-10.3	-8.9	-7.7	-6.5	-5.5	-4.6	-3.9
	6%	-15.2	-13.3	-11.5	-9.9	-8.4	-7.1	-6.0
	8%	-20.0	-17.6	-15.3	-13.2	-11.2	-9.5	-8.1

Table 4: Population expected returns in the Black-Scholes model. The parameter μ is the cum-dividend equity premium, σ is the volatility. These parameters are reported on an annual basis, and expected option returns are monthly percentages.

returns become less negative. For example, for 6% OTM puts with $\mu = 6\%$, EORs change from -39% for $\sigma = 10\%$ to -15% for $\sigma = 20\%$. Thus volatility has a quantitatively large impact and its impact varies across strikes. Unlike the approximately linear relationship between EORs and the equity premium, the relationship between put EORs and volatility is concave. This concavity implies that fully anticipated time-variation in volatility results in more negative expected option returns than that if volatility were constant at the average value.

Expected option returns are extremely difficult to estimate. As a first illustration, the top panel in Figure 2 shows the finite sample distribution for 6% OTM average put returns. The solid vertical line is the observed sample value. The upper panel shows the large variability in average put return estimates: the (5%, 95%) confidence band is -65% to $+28\%$. The figure also shows the marked skewness of the distribution of average monthly option returns, which is expected given the strong positive skewness of purchased put options, and shows why normal approximations are inappropriate.

Table 5 summarizes EORs and p -values corresponding to observed average returns returns for various strikes. Note first that the p -values have increased dramatically relative

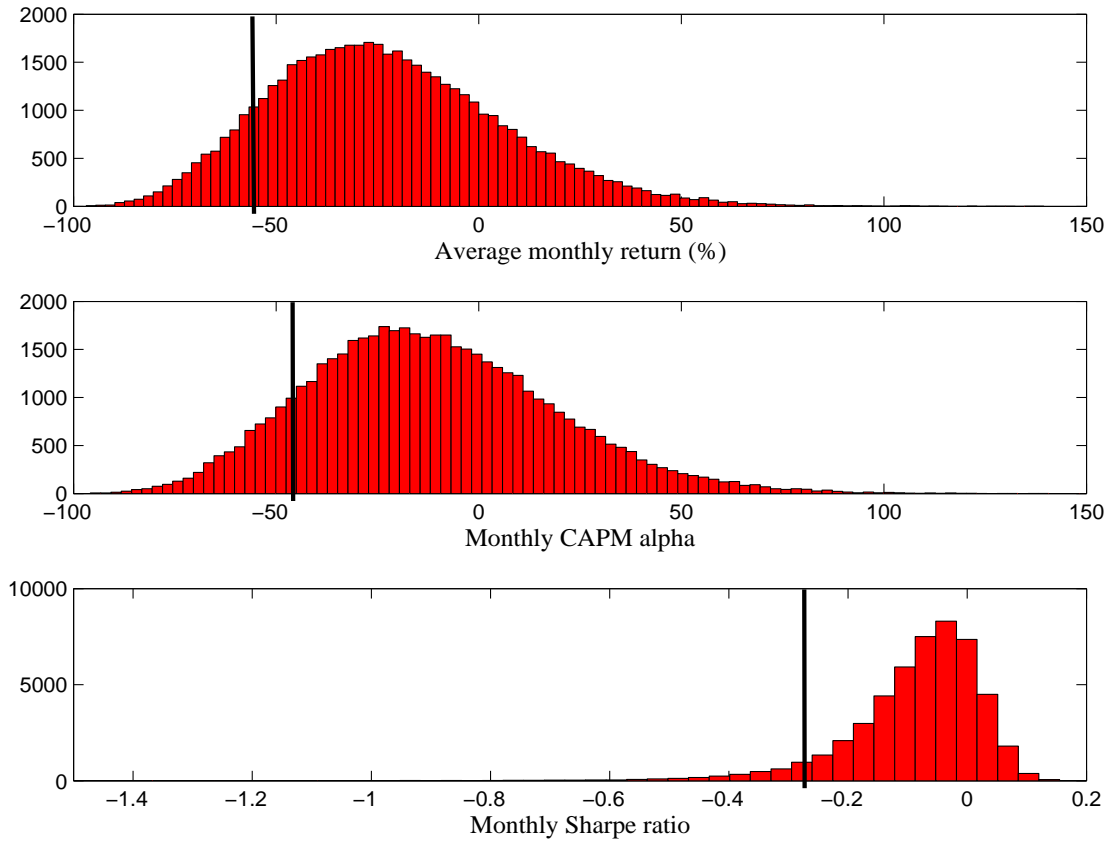


Figure 2: This figure shows histograms of the finite sample distribution of various statistics. The top panel provides the distribution of average 6% OTM put returns, the middle panel 6% OTM put CAPM alphas, and the bottom panel 6% OTM put Sharpe ratios. The solid vertical line is the observed value from the data.

	Moneyiness		0.94	0.96	0.98	1.00
Average returns	Data, %		-56.8	-52.3	-44.7	-29.9
	BS	$E^{\mathbb{P}},\%$	-20.6	-17.6	-14.6	-12.0
		p -value,%	8.1	1.7	0.4	2.2
	SV	$E^{\mathbb{P}},\%$	-25.8	-21.5	-17.5	-13.7
		p -value,%	24.1	9.3	3.0	7.3
CAPM α s	Data, %		-48.3	-44.1	-36.8	-22.5
	BS	$E^{\mathbb{P}},\%$	-17.9	-15.3	-12.7	-10.4
		p -value,%	12.6	2.7	0.3	1.2
	SV	$E^{\mathbb{P}},\%$	-23.6	-19.5	-15.8	-12.4
		p -value,%	39.1	14.1	3.4	8.7
Sharpe ratios	Data		-0.27	-0.29	-0.29	-0.23
	BS	$E^{\mathbb{P}}$	-0.05	-0.07	-0.08	-0.09
		p -value,%	4.9	1.9	1.2	4.0
	SV	$E^{\mathbb{P}}$	-0.04	-0.07	-0.09	-0.10
		p -value,%	21.5	12.0	7.7	14.3

Table 5: This table reports population expected option returns, CAPM α 's, and Sharpe ratios and finite sample distribution p -values for the Black-Scholes (BS) and stochastic volatility (SV) models. We assume that all risk premia (except for the equity premium) are equal to zero.

to Table 2. For example, the p -values using standard t -statistics for the ATM options increase by roughly a factor of 10 and by more than 10,000 for deep OTM put options. This dramatic increase occurs because our bootstrapping procedure anchors null values at those generated by the model (e.g., at negative values, not at zero) and accounts for the large sampling uncertainty in the distribution of average option returns.

Next, average 6% OTM option returns are not statistically different from those generated by the Black-Scholes model, with a p -value of just over 8%. Based only on the Black-Scholes model, we have our first striking conclusion: deep OTM put returns are insignificant, when compared to the Black-Scholes model. This is particularly interesting since the results in the previous literature typically conclude that the deep OTM put options are the most anomalous or mispriced. We arrive at the exact opposite conclusion:

there is no evidence that OTM put returns are mispriced. It is important to note that other strikes are still significant, with p -values below 5%.

Next, consider CAPM alphas, which are reported in the second panel of Table 5. For every strike, the alphas are quite negative and their magnitudes are economically large, ranging from -18% for 6% OTM puts to -10% for ATM puts. Although Black-Scholes is a single-factor model, the alphas are strongly negatively biased in population, which is due to the misspecification discussed at the beginning of Section 2. This shows the fundamental problem that arises when applying linear factor models to nonlinear option returns.

To see the issue more clearly, Figure 3 displays two simulated time series of monthly index and OTM option returns. The regression estimates in the top (bottom) panel correspond to $\alpha = 64\%$ ($\alpha = -51\%$) per month and $\beta = -58$ ($\beta = -19$). The main difference between the two simulations is a *single* large observation in the upper panel, which substantially shifts the constant and intercept estimates obtained by least squares.

More formally, the middle panel of Figure 2 depicts the finite sample distribution of CAPM alphas for 6% OTM puts, and the middle panel of Table 5 provides finite sample p -values for the observed alphas. For the deepest OTM puts, observed CAPM alphas are again insignificantly different from those generated by the Black-Scholes model. For the other strikes, the observed alphas are generally too low to be consistent with the Black-Scholes model, although again the p -values are much larger than those based on asymptotic theory.

Finally, consider Sharpe ratios. The bottom panel of Figure 2 illustrates the extremely skewed finite sample distribution of Sharpe ratios for 6% OTM puts. The third panel of Table 5 reports population Sharpe ratios for put options of various strikes and finite sample p -values. As a comparison, the monthly Sharpe ratio of the underlying index over our sample period is 0.1. The Sharpe ratios are modestly statistically significant for every strike, with p -values between 1% and 5%.

4.2 Stochastic volatility

Next, consider the SV model, which extends Black-Scholes by incorporating randomly fluctuating volatility. We do not assume that the volatility risk is priced, that is, we set $\theta_v^Q = \theta_v^P$. Table 5 provides population average returns, CAPM alphas, and Sharpe ratios for the SV model, as well as p -values.

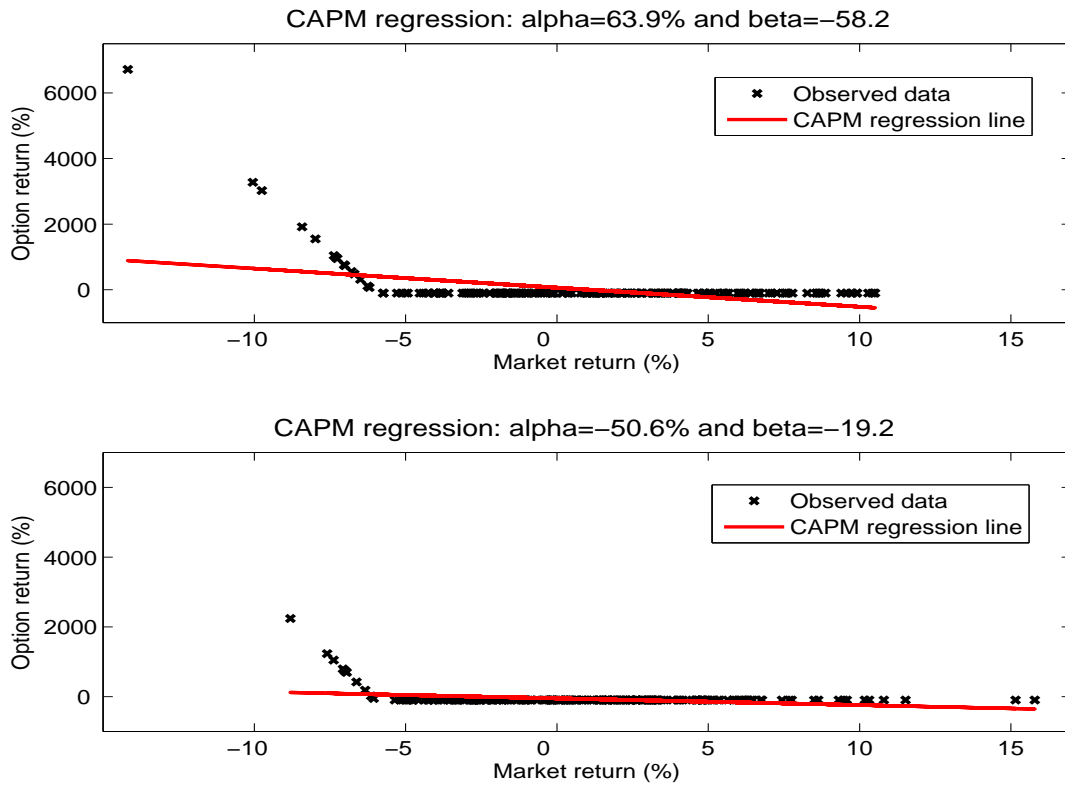


Figure 3: CAPM regressions for 6% OTM put option returns.

Notice first that expected put returns are lower in the SV model. This is due to the fact that EORs are a concave function of volatility, which implies that fluctuations in volatility, even if fully anticipated, decrease expected put returns. Compared to the Black-Scholes model, expected put returns are about 2% lower for ATM strikes and about 5% lower for the deep OTM strikes. While not extremely large, the lower EORs combined with an increased sampling uncertainty generated by changing volatility increase p -values significantly. For deep OTM puts, the p -value is now almost 25%, indicating that roughly one in four simulated sample paths generate average 6% OTM put returns that are *more negative* than those observed in the data. For the other strikes, none of the average returns are significant at the 1% level, and most are not significant at the 5% level.

CAPM alphas for put returns are more negative in population for the SV model than the Black-Scholes model, consistent with the results for expected returns. The observed alphas are all insignificant, with the exception of the 0.98 strike, which has a p -value of about 3%. The results for the Sharpe ratios are even more striking, with none of the strikes statistically significantly different from those generated by the SV model.

4.3 Discussion

The results in the previous section generate a number of new findings and insights regarding relative pricing tests using option returns. In terms of population properties, EORs are quite negative in the Black-Scholes model, and even more so in the stochastic volatility model. The leverage embedded in options magnifies the equity premium and the concavity of EORs as a function of volatility implies that randomly changing volatility increases the absolute value of expected put option returns. Single-factor CAPM-style regressions generate negative CAPM alphas in population, with the SV model generating more negative returns than Black-Scholes. This result is a direct outcome of computing returns of assets with nonlinear payoffs over non-infinitesimal horizons and regressing these returns on index returns. Therefore, extreme care should be taken when interpreting negative alphas from factor model regressions using put returns.

In terms of sampling uncertainty, three results stand out. First, sampling uncertainty is substantial for put returns, so much that the returns for many of the strikes are statistically insignificant. This is especially true for the stochastic volatility model, since randomly changing volatility increases the sampling uncertainty. Second, in terms of statistical efficiency, average returns generally appear to be less noisy than CAPM alphas

or Sharpe ratios. For example, comparing the p -values for the average returns to those for CAPM alphas in the stochastic volatility model, the p -values for CAPM alphas are always larger. This is also generally true when comparing average returns to Sharpe ratios, with the exception of deep OTM strikes. This occurs because the sampling distribution of CAPM alphas is more dispersed, with OLS regressions being very sensitive to outliers. Third, across models and metrics, the most difficult statistics to explain are the 2% OTM put returns. This result is somewhat surprising, since slightly OTM strikes have not been previously identified as particularly difficult to explain.

How do we interpret these results? The BS and SV models are *not* perfect specifications, because they can be rejected in empirical tests using option prices. However, these models do incorporate the major factors driving option returns, and more detailed option model specifications would provide similar features of average monthly option return distributions. The results indicate that average put returns are so noisy that the observed data are not inconsistent with the models. Thus, little can be said when analyzing average put returns, CAPM alphas or Sharpe ratios computed from returns of individual options. If option returns are to be useful, more informative test portfolios must be used.

5 Portfolio-based evidence for option mispricing

This section explores whether returns on option portfolios are more informative about a potential option mispricing than individual option returns. We consider a variety of portfolios including covered puts, which consist of a long put position combined with a long position in the underlying index; ATM straddles, which consist of a long position in an ATM put and an ATM call; crash-neutral straddles, which consists of a long position in an ATM straddle, combined with a short position in one unit of 6% OTM put; put spreads (also known as a crash-neutral puts), which consists of a long position in an ATM put and a short position in a 6% OTM put; and delta-hedged puts, which consist of a long put position with a long position in delta units of the underlying index (because put deltas are negative, the resulting index position is long).

As observed earlier, a large part of the variation in average put returns is driven by the underlying index. All of the above mentioned portfolios mitigate the impact of the level of the index or the tail behavior of the index (e.g., crash-neutral straddles or put spreads). In interpreting the portfolios, the delta-hedged portfolios are the most difficult. Because

Strategy		Delta-hedged puts				ATMS	CNS	PSP
Moneyness		0.94	0.96	0.98	1.00			
Data, %		-1.3	-1.2	-1.0	-0.6	-15.7	-9.9	-21.2
BS	$E^{\mathbb{P}}$, %	0.0	0.1	0.0	0.0	1.1	2.2	-11.1
	p -val, %	0.0	0.0	0.0	0.0	0.0	0.8	12.5
SV	$E^{\mathbb{P}}$, %	-0.3	-0.1	-0.0	0.1	1.4	2.2	-13.1
	p -val, %	0.3	0.0	0.0	0.0	0.0	0.9	17.1

Table 6: Returns on option portfolios. This table reports sample average returns for various put-based portfolios. Population expected returns and finite sample p -values are computed from the Black-Scholes (BS) and stochastic volatility (SV) models. We assume that volatility risk premia are equal to zero. ATMS, CNS and PSP refer to the statistics associated with at-the-money straddles, crash-neutral straddles and put spreads, respectively.

the portfolio weights are either model or data-dependent, deltas vary across models and depend on state variables such as volatility and parameters which need to be estimated. Appendix E provides a detailed accounting of these issues. We use a delta-hedging strategy based on the Black-Scholes model to generate our results (see Appendix E).

In each case, we analyze the returns to the long side to be consistent with the earlier results. In analyzing these positions, we ignore the impact of margin for the short option positions that appear in the crash-neutral straddles and puts. As shown by Santa-Clara and Saretto (2006), margin requirements are substantial for short option positions. Table 6 evaluates expected returns for each of these strategies using the Black-Scholes and SV models from the previous section. CAPM alphas and Sharpe ratios are not reported because they do not add new information, as discussed in the previous section. The table does not include average returns on the covered put positions, since the t -statistics are not significant.

Table 6 shows that the magnitude of the ATM straddle returns is quite large, more than 15% per month, while the magnitude of the delta-hedged returns are much lower, on the order of 1% per month. The corresponding p -values, computed from the finite sample distribution as described in Section 2.2.2, are approximately zero. As expected, the returns on the portfolios are less noisy than for individual option positions. It is interesting to note that put spreads have p -values between 10 and 20 percent and are less significant than individual put returns, at least for strikes that are near-the-money.

ATM straddles have two advantages compared to delta-hedged returns: they can be cleanly interpreted since the portfolio weights are model and data independent and the source of the returns is clear. Straddle returns are related to the wedge between expected volatility under the \mathbb{Q} measure and realized volatility under the \mathbb{P} measure. In our sample, realized monthly volatility is approximately 15% (annualized) and ATM implied volatility 17%.⁸

What is interesting about this gap is that it has not vanished during the latter part of the sample. For example, during the last two one-year periods from July 2003 to July 2004 and July 2004 to July 2005 the gap was 5.3% and 1.9%. This property suggests that the returns cannot be explained solely by factors that have dramatically increased over time such as overall liquidity. The next section investigates two explanations for this gap: estimation risk and jump risk premia.

6 The role of risk premia

In this section, we focus our discussion on straddles and analyzing mechanisms that can generate gaps between \mathbb{P} and \mathbb{Q} measures, although we do report results for delta-hedged returns, crash-neutral straddles, and put spreads. The two mechanisms that we consider are estimation risk and jump risk premia, although we generically refer to the gap as risk premia. Both mechanisms also include an equity premium discussed earlier. We also discuss one other potential explanation, diffusive stochastic volatility risk premia (differences in the drifts of the volatility process under \mathbb{P} and \mathbb{Q}), but find this explanation implausible.

6.1 Differences between \mathbb{P} and \mathbb{Q}

In a purely diffusive stochastic volatility model (i.e., without jumps in volatility), a wedge between expected volatility under \mathbb{P} and \mathbb{Q} can arise from a diffusive volatility risk premium. The simplest version of this assumes that $\theta_v^{\mathbb{Q}} > \theta_v^{\mathbb{P}}$, and has been informally suggested as an explanation for the large straddle returns by, e.g., Coval and Shumway (2001). We argue that this is unlikely to be a main or even a significant driver of the observed straddle returns.

⁸Bakshi and Madan (2006) link this gap to the skewness and kurtosis of the underlying returns via the representative investor's preferences. Chernov (2007) relates this gap to volatility and jump risk premia.

The difference between expected variance under \mathbb{Q} - and \mathbb{P} -measures in the SV model is given by

$$E[V_{t,T}^{\mathbb{Q}}] - E[V_{t,T}^{\mathbb{P}}] = (\theta_v^{\mathbb{Q}} - \theta_v^{\mathbb{P}}) \left(1 + \frac{e^{-\kappa_v^{\mathbb{P}}T} - 1}{\kappa_v^{\mathbb{P}}T} \right),$$

where $E[V_{t,T}^{\mathbb{Q}}]$ denotes the expected average total variance under the measure \mathbb{Q} between time t and $t + T$, assuming continuous observations (see, e.g., Chernov (2006)). Because volatility is highly persistent (i.e., $\kappa_v^{\mathbb{P}}$ is small) and T is small for one-month options, $\theta_v^{\mathbb{Q}}$ needs to be much larger than $\theta_v^{\mathbb{P}}$ to generate a large gap between \mathbb{Q} and \mathbb{P} that is required to obtain the straddle returns consistent with observed data.⁹

Our computations show that $\sqrt{\theta_v^{\mathbb{Q}}} = 22\%$ would generate straddle returns that are statistically insignificant from those observed. However, such a large value of $\theta_v^{\mathbb{Q}}$ implies that the term structure of implied volatilities would be steeply upward sloping on average, which can be rejected based on observed implied volatility term structures, a point discussed by Pan (2002) and Broadie, Chernov, and Johannes (2007). This implies that a large diffusive stochastic volatility risk premium can be rejected as the sole explanation for the short-dated straddle returns. Combined with the previous results, this observation implies that the SV model is incapable of explaining short-dated straddle returns.

A more promising explanation relies on price jumps and to explore this we consider the Bates (1996) and Scott (1997) SVJ model in equations (2.3) and (2.4). From a theoretical perspective, gaps in the \mathbb{P} - and \mathbb{Q} -measure jump parameters are promising because

$$E[V_{t,T}^{\mathbb{Q}}] - E[V_{t,T}^{\mathbb{P}}] = (\theta_v^{\mathbb{Q}} - \theta_v^{\mathbb{P}}) \left(1 + \frac{e^{-\kappa_v^{\mathbb{P}}T} - 1}{\kappa_v^{\mathbb{P}}T} \right) \quad (6.1)$$

$$+ \lambda^{\mathbb{Q}} \left((\mu_z^{\mathbb{Q}})^2 + (\sigma_z^{\mathbb{Q}})^2 \right) - \lambda^{\mathbb{P}} \left((\mu_z^{\mathbb{P}})^2 + (\sigma_z^{\mathbb{P}})^2 \right). \quad (6.2)$$

In contrast to the SV model, differences between the \mathbb{P} - and \mathbb{Q} -measure jump parameters have an impact on expected variance for all maturities and do not depend on slow rates of mean-reversion. The main issue is determining reasonable parameter values for $\lambda^{\mathbb{Q}}, \mu_z^{\mathbb{Q}}$, and $\sigma_z^{\mathbb{Q}}$.

One way to obtain the risk-neutral parameter values is to estimate these parameters from option data, as in Broadie, Chernov, and Johannes (2007). However, this approach is not useful for understanding whether options are mispriced. If option prices are used

⁹We constrain $\kappa_v^{\mathbb{P}} = \kappa_v^{\mathbb{Q}}$. Some authors have found that $\kappa_v^{\mathbb{Q}} < \kappa_v^{\mathbb{P}}$, which implies that $\theta_v^{\mathbb{Q}}$ would need to be even larger to generate a noticeable impact on expected option returns.

to calibrate $\lambda^{\mathbb{Q}}$, $\mu_z^{\mathbb{Q}}$, and $\sigma_z^{\mathbb{P}}$, then the exercise is circular, since one implicitly assumes that options are correctly priced. We take a different approach and consider jump risk premia and estimation risk as plausible explanations for differences between \mathbb{P} - and \mathbb{Q} -measure parameters.

6.2 Factor risk premia

Bates (1988) and Naik and Lee (1990) introduce extensions of the standard lognormal diffusion general equilibrium model incorporating jumps in dividends. These models provide a natural starting point for our analysis. In this application, a particular concern with these models is that, when calibrated to dividends, they lead to well-known equity premium and excess volatility puzzles.¹⁰

Because our empirical exercise seeks to understand the pricing of options given the observed historical behavior of returns, we use the functional forms of the risk correction for the jump parameters, ignoring the general equilibrium implications for equity premium and volatility by fixing these quantities to be consistent with our observed historical data on index returns, 5.4% and 15%, respectively. The risk corrections are given by

$$\lambda^{\mathbb{Q}} = \lambda^{\mathbb{P}} \exp\left(\mu_z^{\mathbb{P}}\gamma + \frac{1}{2}\gamma^2(\sigma_z^{\mathbb{P}})^2\right) \quad (6.3)$$

$$\mu_z^{\mathbb{Q}} = \mu_z^{\mathbb{P}} - \gamma(\sigma_z^{\mathbb{P}})^2, \quad (6.4)$$

where γ is risk aversion, and the \mathbb{P} -measure parameters are those estimated from stock index returns (and not dividend or consumption data) and were discussed earlier. The volatility of jump sizes, $\sigma_z^{\mathbb{P}}$, is the same across both probability measures.

We consider the benchmark case of $\gamma = 10$. This is certainly in the range of values considered to be reasonable in applications. From (6.3) and (6.4), this value generates $\lambda^{\mathbb{Q}}/\lambda^{\mathbb{P}} = 1.65$ and $\mu_z^{\mathbb{P}} - \mu_z^{\mathbb{Q}} = 3.6\%$. The corresponding \mathbb{Q} -parameter values are given in Table 7. We do not consider a stochastic volatility risk premium, $\theta_v^{\mathbb{P}} < \theta_v^{\mathbb{Q}}$, since most standard equilibrium models do not incorporate randomly changing volatility.

¹⁰Benzoni, Collin-Dufresne, and Goldstein (2006) extend the Bansal and Yaron (2004) model to incorporate rare jumps in the latent dividend growth rates. They show that this model can generate a reasonable volatility smile, but they do not analyze the issues of straddle returns, or equivalently, the difference between implied and realized volatility. Their model does not incorporate stochastic volatility.

	$\lambda^{\mathbb{Q}}$	$\mu_z^{\mathbb{Q}}$	$\sigma_z^{\mathbb{Q}}$	$\sqrt{\theta_v^{\mathbb{Q}}}$
Jump risk premia	1.51	-6.85%	$\sigma_z^{\mathbb{P}}$	$\sqrt{\theta_v^{\mathbb{P}}}$
Estimation risk	1.25	-4.96%	6.99%	14.79%

Table 7: \mathbb{Q} -measure parameters for the two scenarios that we explore. In addition, in the estimation risk scenario, we value options with the spot volatility $\sqrt{V_t}$ incremented by 1%.

It is important to note that there are other theories that generate similar gaps between \mathbb{P} and \mathbb{Q} jump parameters. Given the difficulties in estimating the jump parameters, Liu, Pan, and Wang (2004) consider a representative agent who is averse to the uncertainty over jump parameters. Although their base parameters differ, the \mathbb{P} - and \mathbb{Q} -measure gaps they generate for their base parametrization and the “high-uncertainty aversion” case are $\lambda^{\mathbb{Q}}/\lambda^{\mathbb{P}} = 1.96$ and $\mu_z^{\mathbb{P}} - \mu_z^{\mathbb{Q}} = 3.9\%$, which are similar in magnitude to those that we consider.¹¹ We do not have a particular vested interest in the standard risk-aversion explanation vis-a-vis an uncertainty aversion explanation, our only goal is to use a reasonable characterization for the difference between \mathbb{P} - and \mathbb{Q} -measure jump parameters.¹²

6.3 Estimation risk and Peso problems

Another explanation for observed option returns is estimation risk, capturing the idea that parameters and state variables are unobserved and cannot be perfectly estimated from short historical data sets. One argument for why the estimation risk appears in options is provided by Garleanu, Pedersen, and Poteshman (2005). They argue that jumps and discrete trading imply that market makers cannot perfectly hedge, and therefore estimation risk could play an important role and be priced.

In our context, estimation risk arises because it is difficult to estimate the parameters and spot volatility in our models. In particular, jump intensities, parameters of jump size

¹¹Specifically, Liu, Pan, and Wang (2003) assume that $\gamma = 3$, the coefficient of uncertainty aversion $\phi = 20$, and the penalty coefficient $\beta = 0.01$. The \mathbb{P} -measure parameters they use are $\lambda^{\mathbb{P}} = 1/3$, $\mu_z^{\mathbb{P}} = -1\%$ and $\sigma_z^{\mathbb{P}} = 4\%$. We thank Jun Pan for helpful discussions regarding the details of their calibrations.

¹²An additional explanation for gaps between \mathbb{P} and \mathbb{Q} jump parameters is the argument in Garleanu, Pedersen, and Poteshman (2006). Although they do not provide a formal parametric model, they argue that market incompleteness generated by jumps or the inability to trade continuously, combined with exogenous demand pressure, qualitatively implies gaps between realized volatility and implied volatility.

distributions, long-run mean levels of volatility, and volatility mean reversion parameters are all notoriously difficult to estimate. Spot volatility is not observed either. The uncertainty about drift parameters in the stochastic volatility process will have a minor impact on short-dated option returns due to the high persistence of volatility.¹³ The uncertainty in jump parameters can have a significant impact.¹⁴

One way to see the impact of parameter uncertainty is to consider a standard Bayesian setting for learning about the parameters of the jump distribution, assuming jump times and sizes are observed.¹⁵ First, consider uncertainty over the jump mean parameter. Suppose that jump sizes are given by $Z_j = \mu_z + \sigma_z \varepsilon_j$ where $\mu_z \sim \mathcal{N}(\mu_0, \sigma_0^2)$. Then the predictive distribution of Z_{k+1} upon observing k previous jumps is given by

$$p\left(Z_{k+1} \mid \{Z_j\}_{j=1}^k\right) \sim \mathcal{N}\left(\mu_k, \sigma_k^2\right),$$

where

$$\begin{aligned} \mu_k &= w_k \mu_0 + (1 - w_k) \bar{Z}_k, \quad \bar{Z}_k = k^{-1} \sum_{j=1}^k Z_j \\ \sigma_k^2 &= \left(\frac{k}{\sigma_z^2} + \frac{1}{\sigma_0^2}\right)^{-1} + \sigma_z^2, \quad w_k = \frac{\sigma_z^2/k}{\sigma_z^2/k + \sigma_0^2}. \end{aligned}$$

In addition to revising one's beliefs about the location, we also see that $\sigma_k^2 > \sigma_z^2$, implying that estimation risk and learning generates excess volatility. Quantitatively, its impact will be determined by prior beliefs and how many jumps have been observed. In practice, one would expect even more excess volatility, since jump sizes are not perfectly observed.

The impact of uncertainty on σ_z can be even greater. Assuming that μ_z is known, a standard conjugate inverse-gamma prior on the jump variance, $\sigma_z^2 \sim \mathcal{IG}$, implies that the predictive distribution of the jump sizes is t -distributed:

$$p\left(Z_{k+1} - \mu_z \mid \{Z_j\}_{j=1}^k\right) \sim t_\nu,$$

where the degrees of freedom parameter ν depends on the prior parameters and sample size (Zellner, 1971, section 3.2.4). To determine prices, expectations of the form

¹³The argument is similar to the diffusive volatility premium argument in the previous section.

¹⁴Eraker, Johannes, and Polson (2003) provide examples of the estimation uncertainty impact on the implied volatility smiles.

¹⁵Benzoni, Collin-Dufresne, and Goldstein (2006) consider uncertainty over the mean parameters with normal priors. Johannes, Polson, and Stroud (2005) consider sequential learning about jump and stochastic volatility parameters in jump-diffusion setting using historical S&P 500 index returns.

$E(\exp(Z_{k+1})|\mathcal{F}_k)$ will have to be computed. However, if the jump sizes have a t -distribution, this expectation may not exist because the moment generating function of a t -random variable does not exist. Thus, parameter uncertainty can have a substantial impact on the conditional distribution of S_t , as the two examples demonstrate.

Finally, it is also important to consider difficulties in estimating spot volatility. Even with high-frequency data, there are dozens of different methods for estimating volatility, depending on the frequency of data assumed and whether or not jumps are present. Because of this, estimates of V_t are noisy. One could argue that it is possible to estimate V_t from options, but this requires an accurate model and parameter estimates. It is important to note that estimates of V_t differ dramatically across models. In practice, any estimate of V_t is a noisy measure because of all these factors.

To capture the impact of estimation risk, without introducing a formal model for how investors calculate and price estimation risk, we consider the following intuitive approach. We assume that the parameters that we report in Table 1 represent the true data-generating process, that is, these parameters generated the observed S&P 500 index returns over our sample period. However, investors priced options taking into account estimation risk by increasing/decreasing the \mathbb{Q} -measure parameters by one standard deviation from the \mathbb{P} -parameters reported in Table 1. Similarly, to reflect the difficulties in estimating the spot variance V_t , we increase the spot volatility that investors use to value options by 1%. This adjustment is realistic since it implies bid-ask spreads of about 5% for ATM options and 10% for OTM options.¹⁶ The full set of assumed parameter values is reported in the second line of Table 7.

This implementation of estimation risk is also consistent with a “Peso-based” explanation of the deviations between the \mathbb{P} and \mathbb{Q} measures. In this scenario, investors expected a different sample than actually occurred. Potentially, this could mean that fewer jumps were observed or that the realized stochastic volatility path had different characteristics than the observed index return process. Investors priced options accounting for the possibility of more volatility (generated by more jump or higher diffusive volatility), and when these expectations went unfulfilled, put option returns were ex-post quite negative. Thus, the term “Peso problem” could apply to multiple aspects of our model: the parameters of the jump distribution, the jump intensity, parameters of the volatility process, or even

¹⁶Here we compute a bid-ask spread as the difference between an option valued at the theoretical volatility value and an option valued at the adjusted volatility value.

Strategy		Delta-hedged puts				ATMS	CNS
Moneyness		0.94	0.96	0.98	1.00		
Data, %		-1.3	-1.2	-1.0	-0.6	-15.7	-9.9
Jump risk premia	$E^{\mathbb{P}}$, %	-1.3	-1.0	-0.7	-0.4	-11.1	-4.6
	p -val, %	57.0	35.2	10.9	7.7	8.3	6.9
Estimation risk	$E^{\mathbb{P}}$, %	-0.9	-0.7	-0.6	-0.4	-11.0	-7.7
	p -val, %	30.8	18.2	7.6	10.6	12.4	17.5

Table 8: Returns on option portfolios. This table reports sample average returns for various put-based portfolios. Population expected returns and finite sample p -values are computed from the SVJ model for two configurations of \mathbb{Q} -measure parameters. ATMS and CNS refer to the statistics associated with at-the-money and crash-neutral straddles, respectively.

volatility paths. Because of these similarities with the estimation risk, we do not discuss the Peso explanation separately.

6.4 Results

Table 8 reports the results for the jump risk premia and estimation risk explanations, with the parameters given in Table 7. Before discussing the results, it is important to note that the jump risk premia and estimation risk explanations shift the risk-neutral distribution in slightly different ways, since jump risk premia adds more tail mass and estimation risk adds more at-the-money volatility. As discussed earlier, CAPM alphas and Sharpe ratios provide no additional information, even for straddles or delta-hedged returns, and so they are not included.

For both explanations, the p -values are insignificant, for every portfolio and every strike. There are slight differences between estimation risk and jump risk premia, since they impact ATM and OTM options differently. In the risk premium explanation, the lowest p -value occurs for the crash-neutral straddles, but it is still well over 5%. In the estimation risk explanation, the 2% OTM delta-hedged puts generate the lowest p -value, at about 8%.

We interpret these results as follows. While we do not have a formal equilibrium explanation that fits consumption, dividends, stock and option prices, option returns computed from a model with stochastic volatility and jumps which incorporates an equity premium

and jump risk premia and/or estimation risk appear to be consistent with the observed data. Both of these explanations increase risk-neutral volatility relative to observed volatility, and therefore are capable of replicating the observed data. In reality, both estimation risk and jump risk premia are likely important, and therefore an explanation combining aspects of both would be even more palatable. Moreover, as shown by Santa-Clara and Saretto (2005), bid-ask spreads on index options and margin requirements are substantial, and incorporating these would decrease (in absolute value) the observed option returns, making it even easier for models incorporating jump risk premia or estimation risk to explain the observed patterns.

7 Conclusion

In this paper, we propose a new methodology to evaluate the significance of index option returns. To avoid the pitfalls of using individual option returns, CAPM alphas, Sharpe ratios, and asymptotic distributions, we rely on standard option-pricing models to compute analytical expected options returns and to construct finite sample distributions of average option returns using Monte Carlo simulation. When implementing these models, we constrain the equity premium and volatility of stock returns to be equal to the values historically observed, a reasonable assumption when trying to understand option returns (and not equity returns).

We present a number of interesting findings. First, we find that individual put option returns are not particularly informative about option pricing or mispricing. The finite sample distributions are extremely dispersed, due to the difficulty in estimating the equity premium and the highly skewed return distributions generated by put options. In fact, we find that one of the biggest puzzles in the literature, the very large, in absolute value, returns to deep OTM options is, in fact, not inconsistent with the Black-Scholes or Heston stochastic volatility models. Second, we find little added benefit from using CAPM alphas or Sharpe ratios as diagnostic tools because the results are similar to those from average option returns.

Third, we provide evidence that option portfolios, such as straddles or delta-hedged positions, are far more informative, because they are approximately neutral to movements in the underlying. Unlike returns on individual option positions, delta-hedged or straddle returns are shown to be inconsistent with the Black-Scholes and SV models with no risk

premia beyond an equity premium. Finally, we find that option portfolio returns are largely consistent with explanations such as estimation risk or jump risk premia that arise in the context of models with jumps in prices.

We conclude by noting that our results are largely silent on the actual economic sources of the gaps between the \mathbb{P} and \mathbb{Q} measures. It is important to test potential explanations that incorporate investor heterogeneity, discrete trading, model misspecification, or learning. For example, Garleanu, Pedersen, and Poteshman (2005) provide a theoretical model incorporating both investor heterogeneity and discrete trading. It would be interesting to study formal parameterizations of this model to see if it can quantitatively explain the observed straddle returns. We leave these issues for future research.

A Previous research on option returns

Before discussing our approach and results, we provide a brief review of the existing literature analyzing index option returns.¹⁷ The market for index options developed in the mid to late 1980s. The Black-Scholes implied volatility smile indicates that OTM put options are expensive relative to the ATM puts, and the issue is to then determine if these put options are in fact mispriced.

Jackwerth (2000) documents that the risk-neutral distribution computed from S&P 500 index put options exhibits a pronounced negative skew after the crash of 1987. Based on a single factor model, he shows that utility over wealth has convex portions, interpreted as evidence of option mispricing. Investigating this further, Jackwerth (2000) analyzes monthly put trading strategies from 1988 to 1995 and finds that put writing strategies deliver high returns, both in absolute and risk-adjusted levels, with the most likely explanation being option mispricing.

In a related study, Ait-Sahalia, Wang, and Yared (2001) report a discrepancy between the risk-neutral density of S&P 500 index returns implied by the cross-section of options versus the time series of the underlying asset returns. The authors exploit the discrepancy to set up “skewness” and “kurtosis” trade portfolios. Depending on the current relative values of the two implicit densities, the portfolios were long or short a mix of ATM and OTM options. The portfolios were rebalanced every three months. Ait-Sahalia, Wang, and Yared (2001) find that during the period from 1986 to 1996 such strategies would have yielded Sharpe ratios that are two to three times larger than those of the market.

Coval and Shumway (2001) analyze weekly option and straddle returns from 1986 to 1995. They find that put returns are too negative to be consistent with a single-factor model, and that beta-neutral straddles still have significantly negative returns. Importantly, they do not conclude that options are mispriced, but rather that the evidence points toward additional priced risk factors.

Bondarenko (2003) computes monthly returns for S&P 500 index futures options from August 1987 to December 2000. Using a novel test based on equilibrium models, Bon-

¹⁷Prior to the development of markets on index options, a number of articles analyzed option returns on individual securities. These articles, including Merton, Scholes, and Gladstein (1978) and (1982), Gastineau and Madansky (1979), and Bookstaber and Clarke (1985). The focus is largely on returns to various historical trading strategies assuming the Black-Scholes model is correct. Sheikh and Romm (1994) document market microstructure patterns of option returns on individual securities.

darenko finds significantly negative put returns that are inconsistent with single-factor equilibrium models. His test results are robust to risk adjustments, Peso problems, and the underlying equity premium. He concludes that puts are mispriced and that there is a “put pricing anomaly.” Bollen and Whaley (2003) analyze monthly S&P 500 option returns from June 1988 to December 2000 and reach a similar conclusion. Using a unique dataset, they find that OTM put returns were abnormally large over this period, even if delta hedged. Moreover, the pricing of index options is different than individual stock options, which were not overpriced. The results are robust to transaction costs.

Santa-Clara and Saretto (2005) analyze returns on a wide variety of S&P 500 index option portfolios, including covered positions and straddles, in addition to naked option positions. They argue that the returns are implausibly large and statistically significant by any metric. Further, these returns may be difficult for small investors to achieve due to margin requirements and potential margin calls.

Most recently, Jones (2006) analyzes put returns, departing from the literature by considering daily option (as opposed to monthly) returns and a nonlinear multi-factor model. Using data from 1987 to September 2000, Jones finds that deep OTM put options have statistically significant alphas, relative to his factor model. Both in and out-of-sample, simple put-selling strategies deliver attractive Sharpe ratios. He finds that the linear models perform as well or better than nonlinear models. Bates (2006) reviews the evidence on stock index option pricing, and concludes that options do not price risks in a manner consistent with current option-pricing models.

Given the large returns to writing put options, Driessen and Maenhout (2004a) assess the economic implications for optimal portfolio allocation. Using closing prices on the S&P 500 futures index from 1987 to 2001, they estimate expected utility using realized returns. For a wide range of expected and non-expected utility functions, investors optimally short put options, in conjunction with long equity positions. Since this result holds for various utility functions and risk aversion parameters, their finding introduces a serious challenge to explanations of the put-pricing puzzle based on heterogeneous expectations, since a wide range of investors find it optimal to sell puts.

Driessen and Maenhout (2004b) analyze the pricing of jump and volatility risk across multiple countries. Using a linear factor model, they regress ATM straddle and OTM put returns on a number of index and index option based factors. They find that individual national markets have priced jump and volatility risk, but find little evidence of an

international jump or volatility factor that is priced across countries.

B Expected instantaneous option returns

In this appendix, we develop some intuition about the signs, magnitudes, and determinants of instantaneous EORs. First, we apply arguments similar to those used by Black and Scholes to derive their option pricing model for the more general SVJ model. Then we discuss the single-factor Black-Scholes model and its extensions incorporating stochastic volatility and jumps.

B.1 Instantaneous expected excess option returns

The pricing differential equation for a derivative price $f(S_t, V_t)$ in the SVJ model is

$$\begin{aligned} \frac{\partial f}{\partial t} + \frac{\partial f}{\partial S_t} (r - \delta - \lambda^{\mathbb{Q}} \bar{\mu}^{\mathbb{Q}}) S_t + \frac{\partial f}{\partial V_t} \kappa (\theta_v^{\mathbb{Q}} - V_t) + \frac{1}{2} \frac{\partial^2 f}{\partial S_t^2} V_t S_t^2 \\ + \frac{\partial^2 f}{\partial S_t \partial V_t} \rho \sigma_v V_t S_t + \frac{1}{2} \frac{\partial^2 f}{\partial V_t^2} \sigma_v^2 V_t + \lambda^{\mathbb{Q}} E_t^{\mathbb{Q}} [f(S_t e^Z, V_t) - f(S_t, V_t)] = r f, \end{aligned} \quad (\text{B.1})$$

where Z is the jump size and the usual boundary conditions are determined by the type of derivative (e.g., Bates (1996)). We denote the change in the derivative's prices at a jump time, τ_j , as

$$\Delta f_{\tau_j} = f(S_{\tau_j-} e^{Z_j}, V_t) - f(S_{\tau_j-}, V_t)$$

and $F_t = \sum_{j=1}^{N_t} \Delta f_{\tau_j}$.

By Itô's lemma, the dynamics of derivative's price under the measure \mathbb{P} are given by

$$\begin{aligned} df = \left[\frac{\partial f}{\partial t} + \frac{1}{2} \frac{\partial^2 f}{\partial S_t^2} V_t S_t^2 + \frac{\partial^2 f}{\partial S_t \partial V_t} \rho \sigma_v V_t S_t + \frac{1}{2} \frac{\partial^2 f}{\partial V_t^2} \sigma_v^2 V_t \right] dt \\ + \frac{\partial f}{\partial S_t} dS_t^c + \frac{\partial f}{\partial V_t} dV_t + d \left(\sum_{j=1}^{N_t} \Delta f_{\tau_j} \right), \end{aligned} \quad (\text{B.2})$$

where S_t^c is the continuous portion of the index process:

$$\begin{aligned} dS_t^c &= (r + \mu - \delta) S_t dt + S_t \sqrt{V_t} dW_t^s - \lambda^{\mathbb{P}} \bar{\mu}^{\mathbb{P}} S_t dt \\ &= (r + \mu^c - \delta - \lambda^{\mathbb{Q}} \bar{\mu}^{\mathbb{Q}}) S_t dt + S_t \sqrt{V_t} dW_t^s. \end{aligned} \quad (\text{B.3})$$

Substituting the pricing PDE into the drift, we see that

$$\begin{aligned}
df &= \left[-\frac{\partial f}{\partial S_t} (r - \delta - \lambda^{\mathbb{Q}} \bar{\mu}^{\mathbb{Q}}) S_t - \frac{\partial f}{\partial V_t} \kappa(\theta_v^{\mathbb{Q}} - V_t) - \lambda^{\mathbb{Q}} E_t^{\mathbb{Q}} [f(S_t e^Z, V_t) - f(S_t, V_t)] + rf \right] dt \\
&\quad + \frac{\partial f}{\partial S_t} dS_t^c + \frac{\partial f}{\partial V_t} dV_t + dF_t \\
&= (rf - \lambda^{\mathbb{Q}} E_t^{\mathbb{Q}} [f(S_t e^Z, V_t) - f(S_t, V_t)]) dt \\
&\quad + \frac{\partial f}{\partial S_t} [dS_t^c - (r - \delta - \lambda^{\mathbb{Q}} \bar{\mu}^{\mathbb{Q}}) S_t dt] + \frac{\partial f}{\partial V_t} [dV_t - \kappa(\theta_v^{\mathbb{Q}} - V_t)] + dF_t. \tag{B.4}
\end{aligned}$$

From this expression, we can compute instantaneous EORs. Taking \mathbb{P} -measure expectations,

$$\begin{aligned}
\frac{1}{dt} E_t^{\mathbb{P}} [df] &= rf + \frac{\partial f}{\partial S_t} \mu^c S_t + \frac{\partial f}{\partial V_t} \kappa(\theta_v^{\mathbb{P}} - \theta_v^{\mathbb{Q}}) \\
&\quad + \left\{ \lambda^{\mathbb{P}} E_t^{\mathbb{P}} [f(S_t e^Z, V_t) - f(S_t, V_t)] - \lambda^{\mathbb{Q}} E_t^{\mathbb{Q}} [f(S_t e^Z, V_t) - f(S_t, V_t)] \right\}. \tag{B.5}
\end{aligned}$$

Rearranging, instantaneous excess option returns are given by

$$\begin{aligned}
\frac{1}{dt} E_t^{\mathbb{P}} \left[\frac{df(S_t, V_t)}{f(S_t, V_t)} - r dt \right] &= \frac{\partial \log [f(S_t, V_t)]}{\partial \log S_t} \mu^c + \frac{1}{f(S_t, V_t)} \frac{\partial f(S_t, V_t)}{\partial V_t} \kappa(\theta_v^{\mathbb{P}} - \theta_v^{\mathbb{Q}}) \\
&\quad + \frac{\lambda^{\mathbb{P}} E_t^{\mathbb{P}} [f(S_t e^Z, V_t) - f(S_t, V_t)] - \lambda^{\mathbb{Q}} E_t^{\mathbb{Q}} [f(S_t e^Z, V_t) - f(S_t, V_t)]}{f(S_t, V_t)}. \tag{B.6}
\end{aligned}$$

B.2 The Black-Scholes model

In Black-Scholes, the link between instantaneous returns on a derivative, $f(S_t)$, and excess index returns is

$$\frac{df(S_t)}{f(S_t)} = r dt + \frac{S_t}{f(S_t)} \frac{\partial f(S_t)}{\partial S_t} \left[\frac{dS_t}{S_t} - (r - \delta) dt \right].$$

This expression displays two crucial features of the Black-Scholes model. First, instantaneous changes in the derivative's price are *linear* in the index returns, dS_t/S_t . Second, *instantaneous* option returns are conditionally normally distributed. This linearity and normality motivated Black and Scholes to assert a “local” CAPM-style model:

$$\frac{1}{dt} E_t^{\mathbb{P}} \left[\frac{df(S_t)}{f(S_t)} - r dt \right] = \frac{\partial \log [f(S_t)]}{\partial \log (S_t)} \mu.$$

In the Black-Scholes model, this expression shows that EORs are determined by the equity premium and the option's elasticity, which, in turn, are functions primarily of moneyness and volatility.

This instantaneous CAPM is often used to motivate an approximate CAPM model for finite holding period returns,

$$E_t^{\mathbb{P}} \left[\frac{f(S_{t+T}) - f(S_t)}{f(S_t)} - rT \right] \approx \beta_t \mu T,$$

and the model is often tested via an *approximate* linear factor model for option returns

$$\frac{f(S_{t+T}) - f(S_t)}{f(S_t)} = \alpha_T + \beta_t \left(\frac{S_{t+T} - S_t}{S_t} - rT \right) + \varepsilon_{t,T}.$$

As reviewed above, a number of authors use this as a statistical model of returns, and point to findings that $\alpha_T \neq 0$ as evidence of either mispricing or risk premia.

This argument, however, has a serious potential problem since the CAPM does not hold over finite time horizons. Option prices are convex functions of the underlying price, and therefore linear regressions of option returns and underlying returns are generically misspecified. This implies, for example, that α could depend on the parameters $(S_t, K, t, T, \sigma, \mu)$ and is not zero in population. Since the results hold in continuous time, the degree of bias depends on the length of the holding period. Since option returns are highly skewed, the errors $\varepsilon_{t,T}$ are also highly skewed, bringing into question the applicability of ordinary least squares. We show below that even the simple Black-Scholes model generates economically large alphas for put options. These results also bring into question the practice of computing alphas for multi-factor specifications such as the Fama-French model.

B.3 Stochastic volatility and jumps

Consider next the case of Heston's mean-reverting stochastic volatility model (SV). As derived in Appendix B.1, instantaneous realized option returns are driven by both factors,

$$\frac{df(S_t, V_t)}{f(S_t, V_t)} = rdt + \beta_t^s \left[\frac{dS_t}{S_t} - (r - \delta) dt \right] + \beta_t^v [dV_t - \kappa_v^{\mathbb{P}}(\theta_v^{\mathbb{Q}} - V_t)], \quad (\text{B.7})$$

and expected excess returns are given by

$$\frac{1}{dt} E_t^{\mathbb{P}} \left[\frac{df(S_t, V_t)}{f(S_t, V_t)} - rdt \right] = \beta_t^s \mu + \beta_t^v \kappa_v^{\mathbb{P}}(\theta_v^{\mathbb{P}} - \theta_v^{\mathbb{Q}}), \quad (\text{B.8})$$

where

$$\beta_t^s = \frac{\partial \log [f(S_t, V_t)]}{\partial \log S_t} \quad \text{and} \quad \beta_t^v = \frac{\partial \log [f(S_t, V_t)]}{\partial V_t}.$$

Since β_t^v is positive for all options and priced volatility risk implies that $\theta_v^{\mathbb{P}} < \theta_v^{\mathbb{Q}}$, expected put returns are more negative with priced volatility risk.

Equations (B.7) and (B.8) highlight the shortcomings of standard CAPM regressions, even in continuous time. Regressions of excess option returns on excess index returns will potentially generate negative alphas for two reasons. First, if the volatility innovations are omitted then α will be negative to capture the effect of the volatility risk premium. Second, because dS_t/S_t is highly correlated with dV_t , CAPM regressions generate biased estimates of β and α due to omitted variable bias. As in the Black-Scholes case, discretizations will generate biased coefficient estimates.

Next, consider the impact of jumps in prices via Merton's model. Here, the link between option and index returns is far more complicated:

$$\begin{aligned} \frac{df(S_t)}{f(S_t)} = rdt + \frac{\partial \log(f(S_t))}{\partial \log(S_t)} \left[\frac{dS_t^c}{S_t} - (r - \delta - \lambda^{\mathbb{Q}} \bar{\mu}^{\mathbb{Q}}) dt \right] \\ + \left[\frac{f(S_{t-}e^Z) - f(S_{t-})}{f(S_{t-})} \right] - \lambda^{\mathbb{Q}} \frac{E_t^{\mathbb{Q}} [f(S_t e^Z) - f(S_t)]}{f(S_t)} dt, \end{aligned}$$

where dS_t^c denote the continuous portion of the sample path increment and $S_t = S_{t-}e^Z$. The first line is similar to the expressions given earlier, with the caveat that excess index returns contain only the continuous portion of the increment. The second line captures the effect of discrete jumps. Expected returns are given by

$$\frac{1}{dt} E_t^{\mathbb{P}} \left[\frac{df(S_t)}{f(S_t)} - rdt \right] = \beta_t \mu^c + \frac{\lambda^{\mathbb{P}} E_t^{\mathbb{P}} [f(S_t e^Z) - f(S_t)] - \lambda^{\mathbb{Q}} E_t^{\mathbb{Q}} [f(S_t e^Z) - f(S_t)]}{f(S_t)}.$$

Because option prices are convex functions of the underlying, $f(S_{t-}e^Z) - f(S_{t-})$ cannot be linear in the jump size, e^Z , and thus even instantaneous option returns are not linear in index returns. This shows why linear factor models are fundamentally not applicable in models with jumps in prices. For contracts such as put options and standard forms of risk premia (e.g., $\mu_z^{\mathbb{Q}} < \mu_z^{\mathbb{P}}$), $E_t^{\mathbb{P}} [f(S_t e^Z)] < E_t^{\mathbb{Q}} [f(S_t e^Z)]$, which implies that expected put option returns are negatively impacted by any jump size risk premia. As in the case of stochastic volatility, a single-factor CAPM regression, even in continuous time, is inappropriate. Moreover, negative alphas are fully consistent with jump risk premia and are not indicative of mispricing.

C Details of the options dataset

In this appendix, we provide a discussion of major steps taken to construct our options dataset. There are two primary choices to construct a dataset of option prices for multiple strikes: using close prices or by sampling options over a window of time. Due to microstructure concerns with close prices, we followed the latter approach. For each trading day, we select put and call transactions that could be matched within one minute to a futures transaction, typically producing hundreds of matched options-futures transactions. With these matched pairs, we compute Black-Scholes implied volatilities using a binomial tree to account for the early exercise feature of futures options. Broadie, Chernov, and Johannes (2007) show that this produces accurate early exercise adjustments in models with stochastic volatility and jumps in prices.

To reduce the dimension of our dataset and to compute implied volatilities for specific strikes, we fit a piecewise quadratic function to the implied volatilities. This allows us to combine an entire days worth of information and compute implied volatilities for exact moneyness levels. Figure 4 shows a representative day, and Broadie, Chernov, and Johannes (2007) discuss the accuracy of the method. For each month, we select the day that is exactly one month to maturity (28 or 35 calendar days) and compute implied volatilities and option prices for fixed moneyness (in increments of 0.02), measured by strike divided by the underlying.

D Higher frequency data issues

First, it is important to note the advantage of using monthly returns: every month, an option with exactly one-month to maturity exists. This is not the case for weekly returns. This implies that either a single weekly option is used every month (which is severely reduces sample sizes), or weekly returns can be computed by holding a longer-dated option for one week (for example, hold a four-week maturity option for one week, a three-week option for one week, a two-week option for one week, and finally a one-week maturity option until maturity). This strategy has often been utilized in the literature, but it presents an important theoretical complication in this context since weekly return characteristics vary by maturity. A one-week return on a five-week option is theoretically different than a one-week return on a one-week option, generating an “apples and oranges” comparison problem

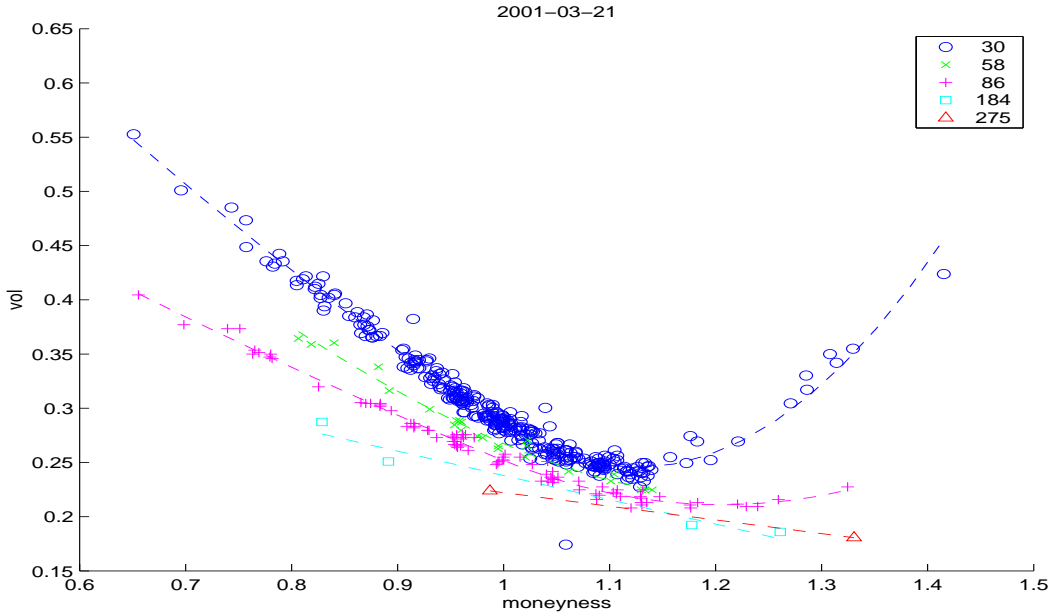


Figure 4: This figure shows representative implied volatility smiles that we construct. Circles represent the actual transactions. The solid line is the interpolated smile.

for weekly returns that is not present in monthly returns.

Second, the option “roll” strategies described above also generates data problems that are particularly acute for deep OTM options. The 6% OTM strike is a relatively actively traded strike for options with one month to maturity. For shorter maturity options, a 6% movement is much less likely, and therefore there is less trading and there are larger bid-ask spreads. In fact, for one-week options, it is often the case that there is virtually no trading in deep OTM strikes. This implies that weekly option returns will generate data availability problems and may also seriously vary in terms of liquidity. For this reason, researchers who construct daily or weekly returns typically allow for moneyness and maturity windows. This strategy adds more data points, but also introduces noise because time to maturity and moneyness vary from one period to another. By using monthly returns, we are able to largely mitigate these issues.

Third, as one considers weekly or daily trading, transaction costs start playing a much more important role than in the monthly buy-and-hold case. The roll strategy is expected to generate inordinate transaction costs. For example, bid-ask spreads are currently on the

order to 3% to 5% of the option price for ATM index options, and are often more than 10% for deep OTM strikes. Moreover, because of liquidity issues mentioned above, transaction costs tend to increase for OTM short-maturity options. We avoid these costs by holding an option throughout the entire month.

Fourth, since our goal is to estimate averages, the insight of Merton (1980) implies that the total time span of data is the key for estimating means rather than the sampling frequency. On the one hand, this would suggest that there is little difference between weekly and monthly returns, at least after they are properly compounded (see, Rubinstein, 1984). On the other hand, intuition from the continuous-time dynamics of an option price (2.2) would suggest that the statistical properties of option returns would improve for higher frequencies.

We analyze this issue by comparing the finite sample distribution of returns in the Black-Scholes and Merton models. We use Monte Carlo simulation to explore whether an increasing frequency of option returns helps in reducing statistical uncertainty of average returns. Simulation analysis is attractive for two reasons. First, we can construct finite sample distributions of average returns and directly compare these distributions under various data sampling scenarios. Second, an idealized simulation setting allows us to abstract from real-life challenges such as transaction costs and additional uncertainty generated by varying maturities and strikes, which are inevitable at frequencies higher than monthly. We can directly focus on the impact of higher frequency.

We simulate from the Black-Scholes and the Merton models many paths, each corresponding to 215 months of observations matching the actual sample. Along each path, we compute average put returns according to two scenarios. First, we replicate the monthly frequency that we use in our dataset. We compute average buy-and-hold returns on a one-month maturity put option. Second, we evaluate a weekly “roll” strategy. We purchase a one-month option and sell it one week later. Then, we take a new option, which has one week less left to maturity and has the same moneyness as the first one. As a result, we construct 860 (four times 215) weekly returns. We compute an average of these returns.

Figure 5 plots the finite sample distributions of average 6% OTM put returns obtained based on our simulation strategy. Consistent with Merton (1980), the statistical properties of average returns are the same for the Black-Scholes model as the frequency increases from monthly to weekly returns, indicating no statistical benefit from using higher frequency returns. Moreover, in the case of the Merton model, higher frequency hurts rather than

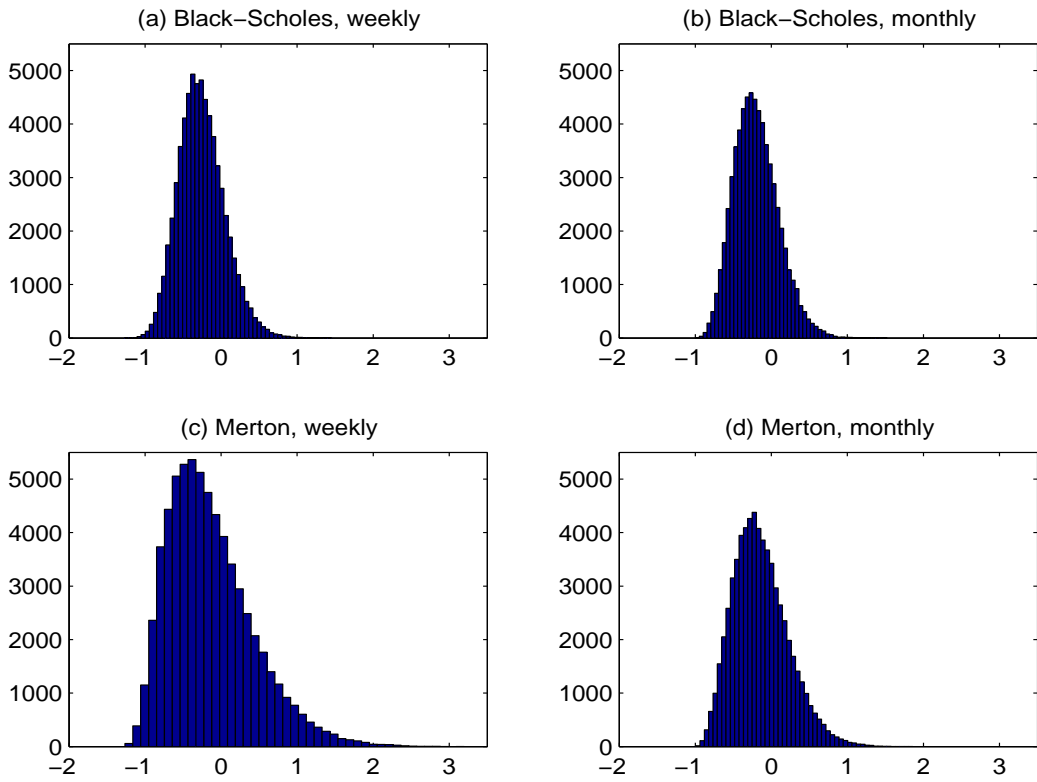


Figure 5: Finite-sample distribution of average option returns.

helps. The distribution of weekly average option returns is more noisy and asymmetric than that of the monthly returns. Therefore, detecting potential option mispricing will be even more challenging at the weekly frequency. The intuition for this result is that the presence of jumps in the underlying asset returns which introduces kurtosis that is more pronounced over short time periods. Therefore, the impact of jumps will be pronounced at high frequencies and will gradually dissipate over longer horizons due to the central limit theorem.

Table 9 complements Figure 5 by providing statistics describing the finite sample distributions of average put returns. In the Black-Scholes case the standard deviation, skewness and kurtosis do not change appreciably from one frequency to another. In contrast, the Merton model implies that weekly frequency leads to dramatic increases in standard devi-

	BSw	BSm	Mw	Mm
Average, %	-25.9	-20.1	-14.3	-13.8
Std Dev, %	30.2	29.2	55.9	36.3
Skew	45.6	51.7	95.6	62.6
Kurt	45.0	35.3	122.4	51.8

Table 9: Put option returns at different frequencies. This table reports statistics describing finite sample distribution of average 6% OTM put returns. BS and M refer to the Black-Scholes and Merton models, respectively. The notation “w” and “m” refers to weekly and monthly frequency of portfolio rebalancing, respectively.

ation, skewness and kurtosis of average returns.

E Delta hedging

In this appendix, we discuss delta-hedged returns and, more generally, returns on strategies with data- or model-based portfolio weights. Delta hedging raises a number of issues that, in our view, make the interpretation of the delta-hedged returns difficult. The main issue is that hedge ratios, or deltas, can be computed in multiple ways.

Here are three approaches to implementing an option delta-hedging strategy. The first uses a formal option pricing model, e.g., the SV model, to compute the required hedging portfolio weights. The second uses a data-based approach that computes the hedge ratios using the shape of the current implied volatility smile (see Bates (2005)). The third, commonly used by practitioners, computes deltas from the Black-Scholes model and substitutes implied volatility for the constant volatility parameter. We discuss each of the approaches in turn.

Model-based hedging requires the knowledge of the spot variance state variable, V_t , and the model parameters. To compute delta-hedged returns using real data we have to estimate V_t in sample. To do this, we require a formal model, which leads to a joint hypothesis issue and introduces estimation noise. Moreover, estimates of spot volatilities and delta-hedged returns are highly sensitive to the model specification and, in particular, to the importance of jumps in prices. For example, Branger and Schlag (2004) show that

delta-hedged errors are not zero if the incorrect model is used or if rebalancing is discrete.

As an alternative, Bates (2005) proposes an elegant model-free technique to establish delta-hedged weights. This approach circumvents the issues mentioned in the previous paragraph. However, the approach assumes that options are priced correctly in the market.¹⁸ This concern is particularly relevant in the context of our paper, because we attempt to evaluate whether options are priced correctly.

The most practical approach is to use a Black-Scholes model delta evaluated at the option's implied volatility. Because of the well-known smile effect in the data, the 6% OTM delta in the Black-Scholes model will be evaluated at an implied volatility that is different from the one used for the ATM delta. This is inconsistent with the Black-Scholes model which assumes a single volatility which is constant across strikes and over time. However, the deltas computed in this manner are similar to the deltas obtained from more elaborate models and estimation procedures.

Finally, delta hedging requires rebalancing, which increases transaction costs and data requirements.¹⁹ Thus, while less attractive from the theoretical perspective, the more practical static delta-hedging strategy should be evaluated. According to this strategy, a delta-hedged position is formed a month prior to an option's maturity and is not changed through the duration of the option contract.

Because static Black-Scholes-based delta-hedging is reasonable and practical, we use it to compute the results in Tables 6 and 8. Thus a consistent delta-hedging strategy is used for generating the results with historical data and the model-based results using the Black-Scholes, SV and SVJ models.

¹⁸Bates (2005) notes: "... while the proposed methodology may be able to infer the deltas ... perceived by the market, that does not mean the market is correct. If options are mispriced, it is probable that the implicit deltas ... are also erroneous."

¹⁹Bollen and Whaley (2004) is the only paper that considers rebalancing. Because of data demands, they take a shortcut and use the volatility at the time the option position was opened and hold this constant until expiration.

References

- Ait-Sahalia, Yacine, Yubo Wang, and Francis Yared, 2001, Do option markets correctly price the probabilities of movement of the underlying asset?, *Journal of Econometrics* 102, 67–110.
- Andersen, Torben, Luca Benzoni, and Jesper Lund, 2002, Towards an empirical foundation for continuous-time equity return models, *Journal of Finance* 57, 1239–1284.
- Bakshi, Gurdip, and Dilip Madan, 2006, A theory of volatility spreads, *Management Science* 52, 1945–1956.
- Bansal, Ravi, and Amir Yaron, 2004, Risks for the long run: A potential resolution of asset pricing puzzles, *Journal of Finance* 59, 1481–1509.
- Bates, David, 1988, Pricing options under jump-diffusion processes, Working paper, University of Pennsylvania.
- , 1996, Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options, *Review of Financial Studies* 9, 69–107.
- , 2005, Hedging the smirk, *Finance Research Letters* 2, 195–200.
- , 2006, The market for crash risk, working paper, University of Iowa.
- Benzoni, Luca, Pierre Collin-Dufresne, and Robert Goldstein, 2005, Can standard preferences explain the prices of out-of-the-money S&P 500 put options, working paper, University of Minnesota.
- Bollen, Nicholas, and Robert Whaley, 2004, Does net buying pressure affect the shape of implied volatility functions?, *Journal of Finance* 59, 711–753.
- Bondarenko, Oleg, 2003, Why are put options so expensive?, Working paper, University of Illinois, Chicago.
- Branger, Nicole, and Christian Schlag, 2004, Can tests based on option hedging errors correctly identify volatility risk premia?, Working paper, Goethe University.
- Broadie, Mark, Mikhail Chernov, and Michael Johannes, 2007, Model specification and risk premia: Evidence from futures options, *Journal of Finance*.

- Cheredito, Patrick, Damir Filipovic, and Robert Kimmel, 2003, Market price of risk specifications for affine models: Theory and evidence, Working Paper, Princeton University.
- Chernov, Mikhail, 2007, On the role of risk premia in volatility forecasting, *Journal of Business and Economic Statistics*.
- , A. Ronald Gallant, Eric Ghysels, and George Tauchen, 2003, Alternative models for stock price dynamics, *Journal of Econometrics* 116, 225–257.
- Driessen, Joost, and Pascal Maenhout, 2004a, A portfolio perspective on option pricing anomalies, Working paper, Insead and University of Amsterdam.
- , 2004b, The world price of jump and volatility risk, Working paper, Insead and University of Amsterdam.
- Eraker, Bjørn, 2004, Do equity prices and volatility jump? Reconciling evidence from spot and option prices, *Journal of Finance* 59, 1367–1403.
- , Michael Johannes, and Nicholas Polson, 2003, The impact of jumps in equity index volatility and returns, *Journal of Finance* 58, 1269–1300.
- Garleanu, Nicolae, Lasse Pedersen, and Allen Poteshman, 2005, Demand-based option pricing, Working paper, University of Pennsylvania.
- Gastineau, G.L., and A. Madansky, 1979, Why simulations are an unreliable test of option strategies, *Financial Analysts Journal* 35, 61–76.
- Johannes, Michael, Nicholas Polson, and Jonathan Stroud, 2005, Sequential parameter estimation in stochastic volatility models with jumps, Working paper, Columbia GSB.
- Jones, Christopher, 2006, A nonlinear factor analysis of S&P 500 index option returns, *Journal of Finance* 41, 2325–2363.
- Merton, Robert, 1980, On estimating the expected return on the market : An exploratory investigation, *Journal of Financial Economics* 8, 323–361.
- Merton, Robert C., Myron S. Scholes, and Mathew L. Gladstein, 1978, The returns and risk of alternative call option portfolio investment strategies, *Journal of Business* 51, 183–242.

- , 1982, The returns and risks of alternative put-option portfolio investment strategies, *Journal of Business* 55, 1–55.
- Naik, Vasanttilak, and Moon Lee, 1990, General equilibrium pricing of options on the market portfolio with discontinuous returns, *Review of Financial Studies* 3, 493–521.
- Pan, Jun, 2002, The jump-risk premia implicit in options: Evidence from an integrated time-series study, *Journal of Financial Economics* 63, 3–50.
- Rubinstein, Mark, 1984, A simple formula for the expected rate of return of an option over a finite holding period, *Journal of Finance* 39, 1503–1509.
- Santa-Clara, Pedro, and Alessio Saretto, 2005, Option strategies: Good deals and margin calls, working paper, UCLA.
- Scott, Louis O., 1997, Pricing stock options in a jump-diffusion model with stochastic volatility and interest rates: Applications of fourier inversion methods, *Mathematical Finance* 7, 413–426.
- Sheikh, Aamir M., and Ehud I. Ronn, 1994, A characterization of the daily and intraday behaviour of returns on options, *Journal of Finance* 49, 557–579.
- Zellner, Arnold, 1971, *An Introduction to Bayesian Inference in Econometrics* (John Wiley & Sons: New York).