Received: 8 April 2017 | Accepted: 12 July 2017
DOI: 10.1002/fut.21865

RESEARCH ARTICLE

WILEY FUTURES MARKETS

Options-based benchmark indices—A review of performance and (in)appropriate measures

Markus Natter

University of Augsburg, Chair of Finance and Banking, Augsburg, Germany

Correspondence
Markus Natter, University of Augsburg, Chair of Finance and Banking, Universitaetsstr. 16, 86159 Augsburg, Germany. Homepage: http://www.wiwi.uni-augsburg.de/bwl/wilkens/team/mitarbeiter/natter_markus/
Email: natter.markus@t-online.de

This paper reviews the performance and profitability of different option strategy benchmark indices provided by the CBOE. Using different performance approaches, I show that performance measurement of these indices is highly complex and sensitive to the model choice. Moreover, this study controls for time-varying delta exposure via linear timing approaches and uses a linear option-factor model that is independent from the portfolio composition. Splitting the sample, I find that outperformance reported by previous studies is mostly driven by limited data. Moreover, the profitability of option strategies for private investors is evaluated based on easily investable investment products.

JEL CLASSIFICATION
G11, G20, G23

1 | INTRODUCTION AND LITERATURE OVERVIEW

In 2002, the CBOE introduced the first option-based benchmark strategy index—the BuyWrite® Index (BXM). Since then, the palette of these indices has experienced dramatic growth. More than 20 strategy indices with different underlying indices and different properties are available on CBOE's homepage.[1] Scientific researchers and practitioners both analyze the performance of these hypothetical portfolios, especially for option-writing indices.[2] This study reviews the performance of strategy benchmark indices using common and novel approaches to measuring profitability. To the best of my knowledge, I am the first to measure the performance of a large number of benchmark indices instead of exclusively one strategy. Moreover, this paper uses a long time horizon as well as different time windows for each strategy index, and introduces novel approaches, for example, conditional factor models, to measure the performance of portfolios containing options.

So far, there are many studies attesting superior performance for CBOE's strategy benchmark indices (e.g., Ungar & Moran, 2009; Whaley, 2002). The first index published was the BXM BuyWrite Index, which is a simple passive covered call strategy that is long the S&P 500 and sells 1-month call options on the underlying index. The original purpose of the BXM was to provide a sufficient benchmark for investors whose portfolios contain options.[3] Whaley (2002) describes the construction of this index and finds more than 20 basis points outperformance on a monthly and risk-adjusted basis compared to the S&P 500.[4] At first glance, this result seems surprising since the BXM strategy theoretically solely invests in the S&P 500 and the risk-free rate, because the

[1]http://www.cboe.com/products/strategy-benchmark-indexes

[2]A collection of contributions on option strategy benchmark indices can be found at http://www.cboe.com/products/strategy-benchmark-indexes/bibliography

[3]Natter et al. (2016) construct a risk-factor derived from BXM's return.

replicating portfolio of a short call option consists of a short position in the respective underlying and a long position in a zero bond. Following a covered call strategy, therefore, means being long the underlying and simultaneously selling some part of the same underlying, whereas the remaining amount of money is invested in a riskless bond. An investment in an index mixed with the risk-free rate should not generate any risk-adjusted outperformance. Nonetheless, there is a vast stream of literature showing the profitability of option writing strategies in the past. Among others, prominent representatives are the studies of Pounds (1978) as well as Bookstaber and Clarke (1981, 1984, 1985). Several studies also find outperformance for CBOE's covered call strategies. Feldman and Roy (2005), for example, report an annualized Jensen's alpha of almost 3% p.a. for the BXM over a 16-year period. Kapadia and Szado (2007) analyze a similar strategy with the Russell 2000 as the underlying index and report an annualized alpha of 2%. Ungar and Moran (2009) measure the performance of the CBOE PutWrite index whose payoff is the analog to a covered call strategy and find an outperformance of over 6% p.a. on a risk-adjusted basis.

This paper contributes to the literature on performance measurement of option-based strategies as follows. (i) So far, most of the recently introduced indices have not been analyzed in detail. (ii) This paper applies and discusses more accurate methods to measure the performance of portfolios containing options by allowing time-varying delta exposure in linear models.[5] (iii) The volatility as the most important determinant of option values is considered in performance measurement via a simple option-factor approach. (iv) By splitting my sample into two separate parts, I show that outperformance documented by previous studies is driven by the first-half (1990–2003) of observations.[6] In addition (v), I examine the profitability of option strategies in different crisis scenarios. While Natter, Rohleder, Schulte, and Wilkens (2016) find superior performance among actively managed equity funds investing in options; I examine whether this result is transferable to passively managed investments and evaluate benefits of direct investments for private investors. Moreover, I show that replicating theoretically calculated indices is non-trivial, as it is associated with costs (vi). Overall, I review and validate previous studies and critically assess their results.

The remainder of the paper is organized as follows. Section 2 introduces the data. Section 3 describes the performance models and section 4 reports the results of the empirical analysis. In section 5, I highlight the possibility to gain exposure toward option strategy benchmark indices. Section 6 concludes.

2 | DATA AND INDEX DESCRIPTION

CBOE's strategy benchmark indexes analyzed in this paper can be distinguished into six subgroups[7]:

2.1 | BuyWrite indexes

A BuyWrite Index or covered call strategy is a passive hypothetical strategy that is long a specific underlying and writes call options on that underlying. Hence, these call options consequently are considered to be covered. The indices of this subgroup differ in their underlyings and as well as in the characteristics of the written call options.

2.2 | PutWrite indexes

A Putwrite Index writes put options on different underlyings, whereas proceeds of the received option premia are invested into riskless T-bills. Hence, the put options are cash-secured and, therefore, also covered. The payoff of this investment strategy is similar to that of a covered call strategy.

[4]The annualized outperformance is, therefore, 2.76% p.a.

[5]Israelov and Nielsen (2015) and Israelov and Klein (2016) develop an approach to decompose the return of option strategies adequately and consider time-varying delta exposure as well. However, their approach requires the knowledge of the exact portfolio composition and data on deltas of actually traded options. The models in this paper shall be translatable to any portfolio return time-series.

[6]Constantinides et al. (2009) find overpricing in S&P 500 options in this period.

[7]A detailed description of the construction of benchmark indices can be found on CBOE's homepage.

2.3 | Combo index

The CBOE S&P 500 Covered Combo Index (CMBO) combines BuyWrite and PutWrite strategies. The index sells 1-month out-of-the-money call options and is simultaneously short out-of-the-money cash-secured put options. The result is a payoff that is similar to those of the individual strategies but steeper at the beginning compared to Buywrite and Putwrite Indexes.

2.4 | Butterfly and condor indexes

The Iron Butterfly Index (BFLY) sells at-the-money call and put options and buys out-of-the-money call and put options. The Iron Condor Index strategy (CNDR) is both short and long out-of-the-money calls and puts with different deltas. The short puts and calls exhibit deltas of about −0.2 and 0.2 and the long options have deltas of −0.05 and 0.05, respectively.

2.5 | Collar indexes

The portfolio of a collar index strategy consists of a long position in the underlying S&P 500 index and a protective out-of-the-money put as well as a written out-of-the-money call. The indexes differ in the characteristics of the options, that is, the portfolios can, for example, be set up at no cost.

2.6 | Put protection index

The Put Protection Index (PPUT) is a passive option benchmark index that follows a simple protective put strategy as proposed by Merton, Scholes, and Gladstein (1978, 1982). The strategy is long the S&P 500 index and buys 5% out-of-the-money put options to limit the potential downside risk of the long position in the equity index.

Historical monthly option benchmark index prices are available with CBOE's website.[8] Since historical data begin and end at different dates, I limit the sample period to February 1990 to the end of 2016 to obtain a comprehensive but comparable sample. Consistent with the literature on performance measurement of option-based benchmark indices, I compare the option strategies to the underlying index and investment products on benchmark indices (see, e.g., Ungar & Moran, 2009; Whaley, 2002). Since I consider only option strategies on the S&P 500, the S&P 500 total return index (SP500TR) serves as the primary benchmark. The data on the S&P 500 total return index are also provided by CBOE, while data on retail investor products stems from Morningstar Direct.

3 | METHODOLOGY

The baseline model for the performance analysis is the CAPM regression and the resulting Jensen Alpha (Jensen, 1968).

$$r_{\mathrm{i,t}} - r_{\mathrm{f}} = \alpha_{\mathrm{i}} + \beta_{\mathrm{i}}(r_{\mathrm{SPTR,t}} - r_{\mathrm{f,t}}) + \varepsilon_{\mathrm{i,t}} \tag{1}$$

where $r_{\mathrm{i,t}}$ is the return of option strategy i and r_{f} is the risk-free rate. Since all of these indices are constructed from the S&P 500 total return index (SPTR), this index is the corresponding market index in the performance regression. α_{i} is Jensen's Alpha and β_{i} is the systematic risk or can be interpreted as the delta of the option strategy, respectively, that is, β_{i} measures the sensitivity of option strategies toward the underlying.

Among others, Lhabitant (2000) and Goetzmann, Ingersoll, Spiegel, and Welch (2007) show that applying risk-adjusted performance measures on option-based investments may lead to potential biases in these figures since portfolios containing options exhibit asymmetric return distributions with fatter tails than the normal. Leland (1999) shows that dynamic strategies consisting of stocks and bonds may bias performance measures, which are grounded in a mean-variance framework. Since the replicating portfolio of an option is a self-financing dynamic strategy consisting of stocks and bonds by definition (Black & Scholes, 1973), even performance measures for buy and hold strategies involving options can be misleading.

Israelov and Nielsen (2015) develop an approach, which accounts for the time-varying character of delta exposure and is, therefore, able to decompose the option index return in single constituents adequately. Their proceeding is very powerful and attributes the performance of an investment in connection with options correctly. However, contrary to this paper, their approach requires deltas for the actual options included in the respective portfolio. The models presented here are supposed to be applicable to

[8]https://www.cboe.com/micro/buywrite/

all portfolios containing options without knowing the exact portfolio holdings positions. To discover the necessity for allowing time-varying systematic risk, consider a simple plain vanilla S&P 500 index call option. The replicating portfolio consists of a long position in delta shares of the S&P 500 and a short position in a zero bond. When the price of the underlying S&P 500 drops dramatically, the share of stocks in the replicating portfolio diminishes as well. It is, therefore, likely that assuming constant betas over time might lead to biased estimators for both performance and systematic risk respectively for the delta of the strategies. Delta indicates the sensitivity of an option or an option strategy to changes in prices of the underlying. Since the underlying of all option strategy benchmark indices is the S&P 500, the beta obtained from a CAPM regression can be interpreted as the option strategy's delta. This leads to the necessity of considering time-varying beta factors in performance models.

Jagannathan and Korajczyk (1986) criticize known timing approaches used in mutual fund research and show both theoretically and empirically that these models may detect spurious timing abilities if portfolios contain options. Conversely, this means that these approaches are able to detect option exposure in portfolios and hence, I employ timing models to measure performance of option strategies. One approach to allow for time-varying betas is similar to the timing model introduced by Treynor and Mazuy (1966) where beta is a linear function of the market return. Using conditional factor models is not necessarily new in finance. Jagannathan and Wang (1996), for example, develop a conditional version of the CAPM and Ferson and Schadt (1996), among others, use conditional factor models to evaluate mutual fund performance. However, to the best of my knowledge, I am the first to apply these models on portfolios containing options.

$$r_{\mathrm{i,t}} - r_{\mathrm{f}} = \alpha_{\mathrm{TM,i}} + \beta_{\mathrm{TM,i,t}}(r_{\mathrm{SPTR,t}} - r_{\mathrm{f,t}}) + \varepsilon_{\mathrm{i,t}} \tag{2}$$

$$\text{with} \quad \beta_{\mathrm{TM,i,t}} = b_{\mathrm{i}} + \gamma_{\mathrm{i}}(r_{\mathrm{SPTR,t}} - r_{\mathrm{f,t}}) \tag{2.1}$$

The slope coefficients of this regression have implications similar to the Greeks of the Black/Scholes/Merton model. $\beta_{\mathrm{TM,i,t}}$ can be interpreted as the index' overall delta. γ_{i} indicates the gamma of the strategies' delta, that is, the sensitivity of delta to changes in the underlying S&P 500.

Another approach including time-varying betas is the Henriksson and Merton (1981) model. This approach is different to the Treynor and Mazuy procedure since beta can only adopt two distinct values. More specifically, the time-varying systematic risk is a function of an option on the market, where 1 is an indicator function that is one if the market excess return is positive and zero otherwise.

$$r_{\mathrm{i,t}} - r_{\mathrm{f}} = \alpha_{\mathrm{HM,i}} + \beta_{\mathrm{HM,i,t}}(r_{\mathrm{SPTR,t}} - r_{\mathrm{f,t}}) + \varepsilon_{\mathrm{i,t}} \tag{3}$$

$$\text{with} \quad \beta_{\mathrm{HM,i,t}} = b_{\mathrm{i}} + \gamma_{\mathrm{if,t}}\, 1_{\{r_{\mathrm{SPTR,t}} - r_{\mathrm{f,t}} > 0\}} \tag{3.1}$$

Options added to portfolios generate asymmetric return distributions with higher moments different from those of the normal distribution. Hence, several new approaches have been proposed by researchers to control for this problem and avoid biases in performance measures.

Leland (1999) argues that dynamic strategies generating skewed return distributions can bias performance measures. In particular, left skewed returns lead to positively biased performance measures. Contrary to that, right skewed returns lead to negatively biased two-dimensional performance figures. To control for skewness, kurtosis and any higher moments of the return distribution, I compute Leland's Alpha for all benchmark indices (see Leland, 1999).

$$\alpha_{\mathrm{L,i}} = E(r_{\mathrm{i}}) - B_{\mathrm{L,i}}[E(r_{\mathrm{SPTR}}) - r_{\mathrm{f}}] - r_{\mathrm{f}} \tag{4}$$

$$\text{where} \quad B_{\mathrm{L,i}} \frac{\mathrm{cov}\Big(r_{\mathrm{i}}; -(1 + r_{\mathrm{SPTR}})\Big)^{-b}}{\mathrm{cov}\Big(r_{\mathrm{SPTR}}; -(1 + r_{\mathrm{SPTR}})\Big)^{-b}} \tag{4.1}$$

$$\text{with} \quad b = \frac{\ln(E(1 + r_{\mathrm{SPTR}})) - \ln(1 + r_{\mathrm{f}})}{\mathrm{Var}\Big(\ln(1 + r_{\mathrm{SPTR}})\Big)} \tag{4.2}$$

Following Whaley (2002), I additionally use his methodology to determine the risk-adjusted performance of option-based strategy benchmark indexes. Whaley's alpha exclusively considers downside risk to calculate systematic risk.

$$\min\left(r_{\mathrm{i,t}} - r_{\mathrm{f,t}}, 0\right) = \alpha_{\mathrm{W,i}} + \beta_{\mathrm{W,i}} \min\left(r_{\mathrm{SPTR,t}} - r_{\mathrm{f,t}}, 0\right) + \varepsilon_{\mathrm{i,t}} \quad (5)$$

In complete markets, option strategies should not be more profitable on a risk-adjusted basis than investments in the underlying. However, there is extant literature that options are richly priced (e.g., Chambers, Foy, Liebner, & Lu, 2014; Constantinides, Jackwerth, & Perrakis, 2009). To disentangle falsely attributed performance not only due to nonlinearities but also resulting from overpriced options, I employ the following option-factor model approach inspired by the work of Coval and Shumway (2001), Broadie, Chernov, and Johannes (2009), and Goyal and Saretto (2009) who, among others, employ models with option-factors.[9] I augment the CAPM regression equation with a straddle-factor to capture non-linear risk exposure that comes from non-linear payoff profiles of options. The option strategy factor is the return time-series of buying a 1-month at-the-money straddle and holding it until expiration. The straddle-factor is then computed as the discrete return of this buy and hold strategy in excess of the risk-free rate.[10]

I calculate option prices using the Black/Scholes/Merton formula using the implied volatility of actually traded S&P 500 index options derived from the Volatility Index (VIX) to obtain a sufficient proxy for option prices.

$$r_{\mathrm{i,t}} - r_{\mathrm{f}} = \alpha_{\mathrm{i}} + \beta_{\mathrm{i,t}}(r_{\mathrm{SPX,t}} - r_{\mathrm{f,t}}) + \beta_{\mathrm{straddle,i}}\ \mathrm{straddle_t} + \varepsilon_{\mathrm{i,t}} \quad (6)$$

$$\text{with}\ \ \mathrm{straddle_t} = \frac{\max\ (\mathrm{SPX_t} - \mathrm{SPX_{t-1}};\ 0) + \max\ (\mathrm{SPX_{t-1}} - \mathrm{SPX_t};\ 0)}{\mathrm{call\_price_{t-1}} + \mathrm{put\_price}_{t-1}} - 1 - r_{\mathrm{f}} \quad (6.1)$$

The straddle-factor model might catch biases arising from asymmetric return distributions as well as from potentially overpriced options.

Many of CBOE's benchmark indices invest in options to hedge portfolio risk or earn premia. The income received from writing options should protect the portfolio from sharp declines in crisis periods. I run an additional performance regression with a crisis dummy variable that equals one if a crisis took place in a given month and zero otherwise. I determine different crisis definitions, for which a detailed description can be found in section 4.5.

$$r_{\mathrm{i,t}} - r_{\mathrm{f}} = \alpha_{\mathrm{i}} + \beta_{\mathrm{M,i}}(r_{\mathrm{SPTR,t}} - r_{\mathrm{f,t}}) + \beta_{\mathrm{C,i}}\ \mathrm{crisis_t} + \beta_{\mathrm{C,M,i}}\left(\mathrm{crisis_t}{*}\left(r_{\mathrm{SPTR,t}} - r_{\mathrm{f}}\right)\right) + \varepsilon_{\mathrm{i,t}} \quad (7)$$

4 | EMPIRICAL RESULTS

4.1 | Descriptive statistics

In a first step, I calculate summary statistics for each group of option strategy indices. Table 1 shows the results.

The discrete returns of the S&P 500 total return index (SPTR) are slightly skewed to the left and a kurtosis in the amount of 4.2162 reflects fatter tails than the normal. The figures of this index serve as standard of comparison for option strategy indices. Option-writing indices, namely BXM, PUT, and CMBO, have average returns similar to the underlying SPTR. However, risk in terms of volatility or semi-deviation, respectively, is considerably lower for option strategies. This leads to a finding in line with the majority of previous studies, namely that these indices outperform on a risk-adjusted basis. The Sharpe Ratio (Sharpe, 1966) is higher for option writing indices than for the underlying index. Other multidimensional performance measures, for example, the Sortino Ratio (Sortino & Price, 1994) or the Stutzer Index (Stutzer, 2000), deliver similar results. The appraisal ratio as introduced by Treynor and Black (1972), which is the constant resulting from a CAPM regression divided by the standard error of the residuals, confirms the superior performance of indices that include short option positions. The payoffs for BXM and PUT exhibit capped upside potentials and so the skewness is more than twice as large in absolute terms compared to the SPTR's (−1.2670 and −1.8886). In contrast to option-writing indices, the PPUT's returns are skewed to the right, since the inherent long put option limits the downside potential of this investment. This is in line with the work of Leland (1999) pointing out that options

[9]Fung and Hsieh (2004) employ a look-back-straddle strategy as a trend-following factor for hedge funds.

[10]Coval and Shumway (2001) were the first to employ at-the-money straddle-factors.

TABLE 1 Summary statistics indices

	SP500 TR	BXM	PUT	CMBO	BFLY	CNDR	CLL	PPUT
Mean (p.a.)	0.1034	0.0888	0.1003	0.0936	0.0628	0.0649	0.0642	0.0680
Vola (p.a.)	0.1443	0.1021	0.0979	0.1060	0.1086	0.0697	0.1033	0.1172
Semi-Vola (p.a.)	0.1605	0.1308	0.1446	0.1343	0.1014	0.1093	0.1034	0.1218
Cum. return	11.0888	8.3794	11.8883	9.5456	3.6131	4.3358	3.8553	4.1576
Skewness	−0.5928	−1.2670	−1.8886	−1.2496	0.0605	−2.0832	−0.1708	−0.3179
Kurtosis	4.2162	7.8561	12.0088	7.4806	2.6015	8.6469	2.7314	3.2375
Sharpe ratio	0.1511	0.1724	0.2139	0.1790	0.0932	0.1542	0.1014	0.0985
Sortino ratio	0.1358	0.1343	0.1444	0.1411	0.0992	0.0974	0.1011	0.0949
Stutzer index	0.1490	0.1664	0.2004	0.1727	0.0934	0.1470	0.1013	0.0982
Omega ratio	1.6475	2.1058	2.9390	2.0762	1.0974	2.8452	1.2587	1.2908
Appraisal ratio		0.0829	0.1580	0.0990	0.0735	0.1123	−0.0902	−0.0929
Max. drawdown	−0.5095	−0.3581	−0.3266	−0.3813	−0.3375	−0.1366	−0.3547	−0.3892

This table shows summary statistics on the S&P 500 total return index as well as on option strategy benchmark indices. Mean, volatility, and semi-volatility are denoted in absolute values on an annualized basis. The sample period spans from 1990 to 2016.

create asymmetric return distributions that are highly skewed to the left or right, respectively. Overall, it is interesting that all indices, which outperform the S&P 500 on a risk-adjusted basis, exhibit left skewed return distributions with fatter tails than the normal. This is first evidence for the necessity to control for higher moments measuring the performance of option strategies.

4.2 | Linear performance models and time-varying betas

Most previous studies regarding option strategy benchmark indices measure the performance using linear performance models. The first problem arising from this approach is that options' delta and risk alter over time. Options can be replicated via stocks and bonds, whereas the replicating portfolio changes every infinitesimal small time step (Black & Scholes, 1973) and thus, beta and delta vary over time. I am the first to consider time-varying delta exposure in a linear regression model via approaches as proposed by Treynor and Mazuy (1966) as well as Henriksson and Merton (1981) when evaluating portfolio containing options. I estimate the models for every index over the whole time period and compare them with a simple CAPM or Jensen's alpha approach, respectively. Table 2 shows the results.

The majority of option indices (BXM, PUT, CMBO, BFLY, and CNDR) outperform the market up to almost 3% points p.a., although this result is only statistically significant for the PUT and CNDR. This is in line with Whaley (2002) and Ungar and

TABLE 2 Performance—Time-varying beta models

	SP500 TR	BXM	PUT	CMBO	BFLY	CNDR	CLL	PPUT
CAPM alpha	0.0000	0.0142	0.0301**	0.0157*	0.0273	0.0256*	−0.0131	−0.0158
CAPM beta	1.0000	0.6180***	0.5592***	0.6616***	0.1003**	0.1502***	0.6531***	0.7389***
TM alpha	0.0000	0.0482***	0.0709***	0.0482***	0.0687***	0.0685***	−0.0430***	−0.0577***
TM beta	1.0000	0.5994***	0.5369***	0.6439***	0.0777	0.1268***	0.6694***	0.7617***
TM gamma	0.0000	−1.5382***	−1.8436***	−1.4682***	−1.8735***	−1.9389***	1.3512***	1.8925**
TM total Perf	0.0000	0.0162	0.0326	0.0176	0.0297	0.0281	−0.0149	−0.0183
HM alpha	0.0000	0.0811***	0.1034***	0.0781***	0.1295***	0.1171***	−0.0637***	−0.0814***
HM beta	1.0000	0.7819***	0.7388***	0.8145***	0.3507***	0.3745***	0.5293***	0.5780***
HM gamma	0.0000	−0.3392***	−0.3717***	−0.3164***	−0.5181***	−0.4642***	0.2564***	0.3329**
HM total Perf	0.0000	0.0177	0.0340	0.0190	0.0327	0.0304	−0.0158	−0.0193

This table shows performance measures for the approaches following Jensen (1968), Treynor and Mazuy (1966) as well as Henriksson and Merton (1981) based on monthly discrete returns. The sample period spans from February 1990 to December 2016. The market index used in all performance regressions is the excess return of the S&P 500 total return index. Performance in terms of alpha is denoted in absolute values p.a. Estimation of standard errors is heteroscedasticity consistent according to White (1980). ***, **, * denote significance of the estimated parameter at the 1%, 5%, and 10% level, respectively.

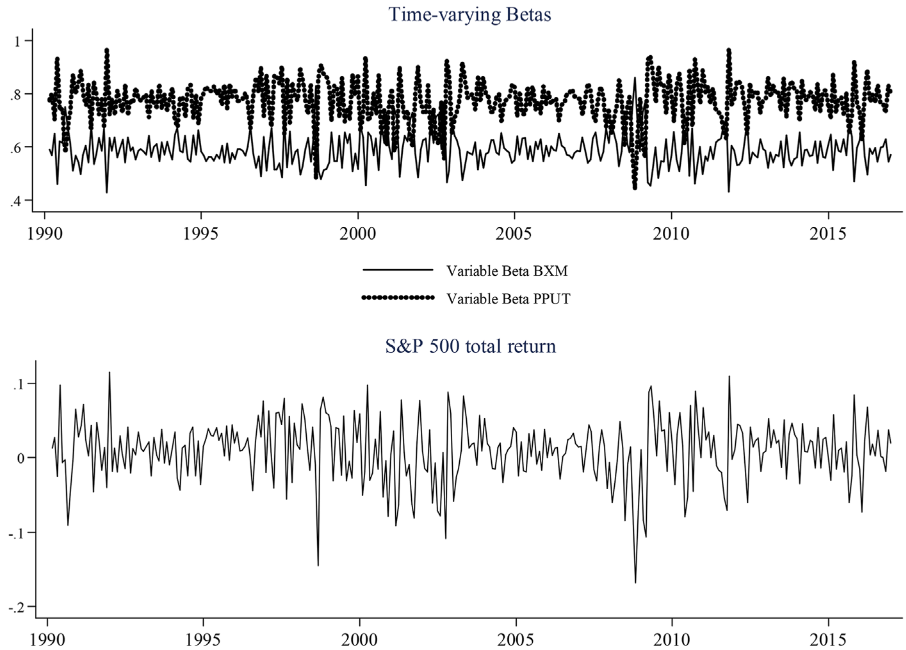

FIGURE 1 Time-varying betas estimated following Treynor/Mazuy (1966). This figure shows time-varying betas for the BXM and PPUT, respectively, estimated via the Treynor and Mazuy (1966) approach. The time period spans from 1990 to 2016

Moran (2009). Since I analyze a longer time period, the results may slightly differ from previous studies. Option strategies should neither exhibit superior nor inferior performance compared to their underlying index after risk-adjustment. The time-varying beta models draw an interesting picture. The gamma of both models is highly significant for all indices and both model specifications, that is, beta depends remarkably on the return of the SPTR.

For further analyses of this phenomenon, the BXM and the PPUT serve to demonstrate the rolling character of time-varying delta and risk exposure of option strategy benchmark indices (Figure 1). Since the payoffs of these two option strategy indices are contrarian, they are especially suitable for displaying the time-varying character of their delta exposure. The BXM sells call options in every month, thus, if the underlying rises and exhibits a positive return, the amount of stocks in the replicating portfolio of the short call option gets more negative and consequently, the BXM's beta drops. The contrary effect can be observed with the PPUT, which is long put options in each month. Long put options are replicated selling stocks short and investing the proceeds in a riskless bond. All else being equal, as the underlying's return increases, the amount of shares sold short diminishes and the beta of the PPUT also increases. Based on the movements shown in Figure 1, I suppose that linear time-varying beta approaches are able to approximate the changing delta exposure character of the strategy indices.

The interpretation of the performance derived from time-varying beta models is non-trivial. Bunnenberg, Rohleder, Scholz, and Wilkens (2017) point out that the constant of these regressions should not be understood as the overall performance. In connection with mutual fund performance, the alpha reflects the selection performance of a fund manager. The overall performance of an investment consists of the selection performance plus the timing performance. Passive option strategy indices should exhibit neither selection nor timing. However, it is possible to compute the total performance for any investment analytically and, therefore, I follow Bunnenberg et al. (2017) to determine option indices' total performance. For both timing model approaches the total performance is not clearly distinguishable from the total performance measured by Jensen's alpha. While the change in alphas is 54 basis points for the Henriksson and Merton total performance of the BFLY compared to its Jensen's alpha, all other figures show smaller alterations without losing their original sign and economic significance. Modeling time-varying delta exposure seems to be necessary but does not enhance or change the measured performance figures remarkably. A possible reason for these similar results could be that the beta of option strategies is not a linear function of the underlying's return. Israelov and Nielsen (2015) show theoretically that the functional relation proposed by Treynor and Mazuy (1966) and Henriksson and Merton (1984) is not feasible for portfolios containing options.

TABLE 3 Performance—Considering higher moments

	SP500 TR	BXM	PUT	CMBO	BFLY	CNDR	CLL	PPUT
CAPM alpha	0.0000	0.0142	0.0301**	0.0157*	0.0273	0.0256*	−0.0131	−0.0158
CAPM beta	1.0000	0.6180***	0.5592***	0.6616***	0.1003**	0.1502***	0.6531***	0.7389***
Leland's alpha	0.0000	0.0226	0.0397	0.0230	0.0496	0.0466	−0.0015	−0.0056
Leland's beta	1.0000	0.6396	0.5858	0.6819	0.1277	0.1768	0.6346	0.7116
Whaley's alpha	0.0000	0.0156***	0.0208***	0.0154***	−0.1036***	−0.0277***	−0.0350***	−0.0393***
Whaley's beta	1.0000	0.7027***	0.6465***	0.7348***	0.2146***	0.2692***	0.5843***	0.6520***

This table shows performance measures for the approaches following Jensen (1968), Leland (1999) as well as Whaley (2002) based on monthly discrete returns. The sample period spans from February 1990 to December 2016. The market index used in all performance calculations is the excess return of the S&P 500 total return index. Performance in terms of alpha is denoted in absolute values p.a. Estimation of standard errors is heteroscedasticity consistent according to White (1980). ***, **, * denote significance of the estimated parameter at the 1%, 5%, and 10% level, respectively.

4.3 | Controlling for higher moments

In section 3, I show that controlling for higher moments is theoretically essential. Therefore, I employ Leland's approach for all option indices in my sample over the whole time period. Moreover, Whaley (2002) published the first study examining the BXM BuyWrite Index and proposes an approach that only considers downside risk in returns. Consequently, I estimate Equation (5) for the option strategy benchmark indices. Table 3 displays the results.

As Leland's alpha is not obtained from a linear regression, there is no test on significance reported in Table 3. An interesting finding is that the alphas obtained by both Leland's and Whaley's models get even more pronounced instead of changing the direction due to left skewed returns. For the BXM, Leland's alpha rises from 1.42% (Jensen's alpha) up to 2.26%, although the skewness is highly negative (−1.2670). The estimated parameter for the PPUT are in line with Leland's theory, since this strategy generates less negatively skewed returns (−0.3179) compared to the S&P 500 and hence, the alpha increases, respectively, becomes less negative. The same result is observable for the CLL whose skewness is even more positive.

Alphas measured via Whaley's model are now throughout statistically significnant. Alpha for the BFLY changes dramatically in sign from 2.73% to −10.36%, whereas all other figures point in the same direction. This result shows the sensibility of measured performance and the dependence of the chosen model. However, I observe that controlling for higher moments in performance models does not change results significantly similar to the outcomes for time-varying beta models.

TABLE 4 Performance—Straddle-factor model

	SP500 TR	BXM	PUT	CMBO	BFLY	CNDR	CLL	PPUT
CAPM alpha	0.0000	0.0142	0.0301**	0.0157*	0.0273	0.0256*	−0.0131	−0.0158
CAPM beta	1.0000	0.6180***	0.5592***	0.6616***	0.1003**	0.1502***	0.6531***	0.7389***
Panel A								
Alpha (OF)	0.0000	−0.0264**	−0.0130	−0.0219*	−0.0442*	−0.0246	0.0137	0.0159
Beta (OF)	1.0000	0.6153***	0.5563***	0.6592***	0.0956**	0.1469***	0.6549***	0.7409***
Straddle (OF)	0.0000	−0.0116***	−0.0123***	−0.0107***	−0.0204***	−0.0144***	0.0077***	0.0091***
Panel B								
Alpha (OF2)	0.0000	−0.0196	−0.0066	−0.0141	−0.0295	−0.0142	0.0135	0.0178
Beta (OF2)	1.0000	0.5878***	0.5307***	0.6278***	0.0366	0.1050**	0.6558***	0.7332***
Straddle (OF2)	0.0000	−0.0107***	−0.0114***	−0.0097***	−0.0184***	−0.0129***	0.0076***	0.0093***
VIX (OF2)	0.0000	−0.0089	−0.0083	−0.0102*	−0.0192*	−0.0136*	0.0003	−0.0025

This table shows performance measures for the approaches following Jensen (1968) and for my option-factor approach in the spirit of Coval and Shumway (2001), Broadie et al. (2009) as well as Goyal and Saretto (2009) based on monthly discrete returns. The sample period spans from February 1990 to December 2016. The market index used in all performance regressions is the excess return of the S&P 500 total return index. The straddle-factor is the return of a simple buy-and-hold strategy in excess of the risk-free rate: in month t open a long call and put position and hold it 1 month until expiration and repeat this procedure every month. Panel A displays the results for a CAPM regression augmented with the straddle-factor, whereas Panel B contains outcomes for the same model with the return of the VIX as an additional risk-factor. Performance in terms of alpha is denoted in absolute values p.a. Estimation of standard errors is heteroscedasticity consistent according to White (1980). ***, **, * denote significance of the estimated parameter at the 1%, 5%, and 10% level, respectively.

4.4 | Option-factor models

As introduced in section 3, I employ an option-strategy model to measure the performance of option strategies. Table 4 displays the outcome of the option-factor model regressions. The first thing to mention is that the augmented straddle option-factor seems to catch option exposure in option indices adequately (Panel A). Indices systematically selling options show negative loadings on the return of the straddle strategy and in contrast, the coefficient for the PPUT index, which is long put options, is significantly positive. Since the straddle-factor is constructed from both call and put options, I assume that this factor is able to detect the correct overall exposure in a portfolio investing (partly) in options with different long and short positions.[11]

As the additional factor is almost orthogonal to the SPTR,[12] the market beta does not change in magnitude and thus, multicollinearity seems not to be an issue in my test setting. However, performance in terms of alpha changes dramatically in sign for all indices. For BXM, CMBO and BFLY the underperformance of up to almost 4.5% points per year is not only statistically, but also economically significant. The change in sign deserves more attention. According to Leland (1999) left skewed returns generate positively biased alphas and vice versa. If the option-factor model adequately catches non-linearities, the alphas should be pulled into the opposite direction. Again, the BXM and the PPUT serve to illustrate this mechanism: Jensen's alpha for the BXM is positive and the performance measured with the augmented straddle-factor is negative. In contrast, the alpha of the PPUT, which exhibits returns that are skewed to the right, changes from a negative to a positive value. The problem of this model specification is that I am not able to distinguish between biases that come from asymmetric return distributions and such arising from overpriced options. This might be a reason why the alphas of some indices become statistically significant, that is, they could be under- or overestimated.

Another main driver of the profitability of option strategies is the overall market price level of traded options. It seems plausible that in times when the price of options is generally high, selling options should be more lucrative. The straddle-factor indirectly reflects the price level due to the implied volatility that is used to compute actual traded option prices. For a deeper understanding of the results, I augment Equation (4) with the return of the VIX in excess of the risk-free rate. Every alpha is not statistically indistinguishable from zero anymore. The VIX-factor shows weak statistical significance, especially for some indices, which exhibited a statistically significant alpha in the previous regression setting.

However, one should interpret the results displayed in Panel B of Table 4 carefully. For an adequate interpretation of alpha as risk-adjusted performance as a return term, the risk-factors must be directly investable. This is true for the straddle-factor since it is a simple and repeated passive buy-and-hold strategy, which can be invested at relatively low cost. On the contrary, a direct investment into the Volatility Index (VIX) is not possible. There are indeed some investment vehicles, which allow exposure that is highly correlated with the VIX. Futures and options on the VIX are traded at CBOE and investment banks offer ETFs and ETNs on these futures.[13] All in all, there is no such investment product, which exactly reflects the development of CBOE's VIX and hence, the alpha of this performance model cannot be directly interpreted as performance denoted in percentage points.

4.5 | Crises analysis

As shown in summary statistics, option strategy benchmark indices exhibit lower risk in terms of volatility. The indices are said to have an inherent protection against market declines. Indices that sell options have a cushion in the amount of the option premium that prevents the portfolios from potential losses. Indices that buy options, especially puts (e.g., the PPUT), should also be limited in losses. Schulte and Stamos (2015) find abnormally high returns for option strategies in the recent financial crisis (2008–2010) in the long run. Therefore, it seems reasonable to test whether these strategies strategies prevent the investor from experiencing losses in crises periods.

I determine five different crises scenarios: crisis scenario (i) is defined following Chalmers, Kaul, and Phillips (2013). Scenario (ii) defines all times as crisis when the implied volatility is extremely high, that is, the VIX exceeds the 75th percentile. Defining scenario (iii), I obtain business cycle data that indicate recessions in the United States from NBER and St. Louis Fed, respectively.[14] The first three scenarios reflect acknowledged crises definitions. Finally, I specify two different crises scenarios with respect to the S&P500 total returns. Scenario (iv) considers all points in time as crisis period when the return of the SPTR

---

[11]In contrast, Agarwal and Naik (2004) only use separate call and put option-factors in their work on hedge funds.

[12]The linear correlation between the excess return of the SPTR and the straddle-factor is −0.02.

[13]An example for an ETN on the VIX is offered by Barclays Capital iPath®: iPath® S&P 500 VIX Short-Term FuturesTM ETN (VXX) http://www.ipathetn.com/US/16/en/details.app?instrumentId=259118. This product demands a fee of 0.89% p.a., which is not included in the VIX-factor used in the performance model above.

[14]Data are freely available at St. Louis Fed's homepage: https://fred.stlouisfed.org/series/USREC

TABLE 5 Crisis analysis

	BXM	PUT	CMBO	BFLY	CNDR	CLL	PPUT
Scenario (i)							
Alpha	0.0209**	0.0366***	0.0239***	0.0295	0.0281*	−0.0247***	−0.0250***
Beta	0.5714***	0.5002***	0.6133***	0.0871	0.1126**	0.7193***	0.8208***
Crisis	−0.0138	−0.0030	−0.0232	−0.0059	0.0100	0.0338	0.0051
Interaction	0.1436**	0.1857**	0.1470**	0.0402	0.1211	−0.2004***	−0.2577**
Scenario (ii)							
Alpha	0.0344***	0.0538***	0.0327***	0.0576**	0.0612***	−0.0373***	−0.0388***
Beta	0.4761***	0.4041***	0.5412***	−0.0868	−0.0129	0.7757***	0.8898***
Crisis	0.0003	−0.0062	0.0008	−0.0151	−0.0527	0.0287	0.0058
Interaction	0.2424***	0.2624***	0.2060***	0.3136***	0.2574***	−0.1977***	−0.2554***
Scenario (iii)							
Alpha	0.0228**	0.0402***	0.0232***	0.0369	0.0310**	−0.0253***	−0.0261***
Beta	0.5750***	0.5040***	0.6245***	0.0522	0.1114***	0.7127***	0.8048***
Crisis	−0.0310	−0.0289	−0.0277	−0.0335	−0.0039	0.0445	0.0180
Interaction	0.1560**	0.2041*	0.1346*	0.1750*	0.1490*	−0.2158**	−0.2496*
Scenario (iv)							
Alpha	0.0507***	0.0636***	0.0508***	0.0894***	0.0829***	−0.0277***	−0.0415***
Beta	0.4938***	0.4339***	0.5451***	−0.0994	−0.0312	0.7290***	0.8461***
Crisis	0.1304***	0.1838***	0.1086***	0.1565*	0.1282***	−0.1995***	−0.2042***
Interaction	0.4159***	0.4874***	0.3724***	0.5997***	0.5267***	−0.4096***	−0.4782***
Scenario (v)							
Alpha	0.0419***	0.0547***	0.0492***	0.0792**	0.0810***	−0.0101	−0.0245***
Beta	0.5090***	0.4497***	0.5470***	−0.0825	−0.0285	0.6950***	0.8145***
Crisis	0.1109***	0.1384***	0.0816***	0.1425**	0.1021***	−0.1527***	−0.1620***
Interaction	0.3697***	0.4102***	0.3388***	0.5579***	0.4922***	−0.2998***	−0.3783***

This table shows performance measures for different crisis scenarios following Jensen (1968). The sample period spans from February 1990 to December 2016. The market index used in all performance regressions is the excess return of the S&P 500 total return index. Scenario (i) is defined as in Chalmers et al. (2013); scenario (ii) defines times when the implied volatility measured by the VIX exceeds the 75th percentile; scenario (iii) reflects business cycles obtained from NBER and St. Louis Fed. The crisis dummy in scenario (iv) is one if the SPTR's return is below the 25th percentile and in scenario (v), the binary crises variable takes on the value one if the SPTR's return is negative. Performance in terms of alpha is denoted in absolute values p.a. Estimation of standard errors is heteroscedasticity consistent according to White (1980). ***, **, * denote significance of the estimated parameter at the 1%, 5%, and 10% level, respectively.

falls below the 25th percentile, whereas scenario (v) takes all times into account where the SPTR's return was negative. For every crisis scenario, I run the model given in Equation (6). Table 5 shows the result of this regression model.

As Equation (6) is not a performance model regression in the classical meaning, the interpretation of the parameters is not straightforward. Alpha itself shall not be seen as performance in terms of return denoted in percentage points, as parts of the factors in the model are not investable. Instead, I focus on the analysis of the crisis dummy variable and the interaction with the SPTR's return. For scenarios (i), (ii), as well as (iii), the crisis dummy is consistently indistinguishable from zero. As this coefficient may be interpreted as additional performance effect during crisis periods, it seems that option strategy indices do not provide a shelter from losses in turbulent times in the classical meaning. However, the last two scenarios that are determined solely by the SPTR's returns exceeding a certain threshold, exhibit crisis dummy coefficients that are throughout statistically significant. For the BXM, PUT, CMBO, BFLY as well as CNDR, the estimated parameters are positive. According to this result, I conclude that these option strategies are indeed able to protect an investor from experiencing drawdowns due to crashes in the underlying S&P 500. On the other hand, one can observe significantly negative coefficients for the CLL and PPUT, which means that these indices tend to perform weaker in crises. The results attained from this analysis could be caused by the fact that indices selling options earn premia that serve as a cushion against strong market declines. On the contrary, indices that are mainly long options suffer from paying options premia and, therefore, underperform in crises. A possible reason for this phenomenon are

TABLE 6 Statistics for the split sample

Panel A: 1990–2003								
	SP500 TR	**BXM**	**PUT**	**CMBO**	**BFLY**	**CNDR**	**CLL**	**PPUT**
Mean (p.a.)	0.1211	0.1188	0.1248	0.1220	0.1094	0.0945	0.0914	0.0817
Vola (p.a.)	0.1495	0.1018	0.0929	0.1046	0.1037	0.0697	0.1099	0.1254
Sem-Vola (p.a.)	0.1607	0.1289	0.1410	0.1303	0.0930	0.1152	0.1065	0.1283
Cum. return	3.5854	3.8232	4.2991	4.0199	3.2309	2.5796	2.2694	1.7884
Skewness	−0.4768	−1.2173	−1.8530	−1.1819	−0.0453	−2.1371	−0.0876	−0.2187
Kurtosis	3.5349	6.6433	9.9957	6.6338	2.2082	8.8091	2.5797	2.9031
Sharpe ratio	0.1511	0.2157	0.2553	0.2187	0.1853	0.2141	0.1274	0.0894
Sortino ratio	0.1404	0.1700	0.1676	0.1752	0.2063	0.1291	0.1314	0.0873
Stutzer index	0.1498	0.2071	0.2377	0.2101	0.1858	0.2005	0.1276	0.0894
Omega ratio	1.5692	2.2115	3.2821	2.1509	1.1974	3.3947	1.2877	1.2568
Appraisal ratio		0.1707	0.2241	0.1873	0.1715	0.1732	−0.0210	−0.1459
Max. drawdown	−0.4473	−0.3019	−0.2900	−0.3210	−0.1193	−0.1332	−0.2070	−0.3453
CAPM alpha	0.0000	0.0296**	0.0430**	0.0301**	0.0615**	0.0394**	−0.0034	−0.0225
CAPM beta	1.0000	0.5928***	0.4979***	0.6267***	0.0642	0.1545***	0.6640***	0.7846***
Alpha (OF)	0.0000	−0.0073	0.0029	−0.0036	0.0014	−0.0070	0.0246*	0.0041
Beta (OF)	1.0000	0.5987***	0.5043***	0.6320***	0.0737	0.1619***	0.6595***	0.7803***
Straddle (OF)	0.0000	−0.0121***	−0.0132***	−0.0111***	−0.0198***	−0.0153***	0.0092***	0.0088***
TM alpha	0.0000	0.0728***	0.0912***	0.0715***	0.1286***	0.0961***	−0.0383***	−0.0542***
TM beta	1.0000	0.5789***	0.4824***	0.6134***	0.0426	0.1363***	0.6752***	0.7947***
TM gamma	0.0000	−1.8601***	−2.0732***	−1.7796***	−2.8876***	−2.4377***	1.4997***	1.3628***
HM alpha	0.0000	0.1054***	0.1274***	0.1017***	0.1852***	0.1501***	−0.0693***	−0.0763***
HM beta	1.0000	0.7742***	0.6998***	0.7979***	0.3600***	0.4193***	0.5066***	0.6560***
HM gamma	0.0000	−0.3652***	−0.4065***	−0.3447***	−0.5956***	−0.5330***	0.3170***	0.2589**
Panel B: 2004–2016								
	SP500 TR	**BXM**	**PUT**	**CMBO**	**BFLY**	**CNDR**	**CLL**	**PPUT**
Mean (p.a.)	0.0845	0.0566	0.0740	0.0631	0.0129	0.0332	0.0350	0.0532
Vola (p.a.)	0.1388	0.1020	0.1028	0.1072	0.1122	0.0689	0.0954	0.1081
Semi-Vola (p.a.)	0.1602	0.1326	0.1478	0.1382	0.1089	0.1042	0.1000	0.1142
Cum. return	1.6364	0.9446	1.4322	1.1007	0.0903	0.4906	0.4851	0.8497
Skewness	−0.7466	−1.3252	−1.8903	−1.3142	0.1890	−2.0546	−0.3419	−0.4719
Kurtosis	5.1307	9.1387	13.1610	8.2667	2.9512	8.6021	2.8344	3.6895
Sharpe ratio	0.1507	0.1261	0.1742	0.1375	0.0025	0.0894	0.0697	0.1101
Sortino ratio	0.1309	0.0972	0.1213	0.1069	0.0026	0.0589	0.0666	0.1045
Stutzer index	0.1483	0.1229	0.1653	0.1338	0.0025	0.0871	0.0697	0.1095
Omega ratio	1.7368	2.0000	2.6279	2.0000	1.0000	2.3913	1.2286	1.3284
Appraisal ratio		−0.0147	0.0871	0.0001	−0.0248	0.0472	−0.1923	−0.0453
Max. drawdown	−0.5095	−0.3581	−0.3266	−0.3813	−0.3375	−0.1366	−0.3547	−0.3892
CAPM alpha	0.0000	−0.0025	0.0160	0.0000	−0.0095	0.0108	−0.0234*	−0.0082
CAPM beta	1.0000	0.6487***	0.6346***	0.7045***	0.1441**	0.1445**	0.6394***	0.6825***
Alpha (OF)	0.0000	−0.0477***	−0.0270	−0.0417***	−0.1002***	−0.0456**	−0.0033	0.0281
Beta (OF)	1.0000	0.6359***	0.6224***	0.6927***	0.1185*	0.1286***	0.6451***	0.6928***
Straddle (OF)	0.0000	−0.0114***	−0.0108***	−0.0105***	−0.0228***	−0.0141***	0.0050***	0.0091***
TM alpha	0.0000	0.0236	0.0483***	0.0242*	0.0108	0.0433**	−0.0482***	−0.0567***

(Continues)

TABLE 6 (Continued)

Panel B: 2004–2016								
	SP500 TR	**BXM**	**PUT**	**CMBO**	**BFLY**	**CNDR**	**CLL**	**PPUT**
TM beta	1.0000	0.6266***	0.6072***	0.6840***	0.1269	0.1169**	0.6605***	0.7236***
TM gamma	0.0000	−1.2417**	−1.5424**	−1.1544***	−0.9667	−1.5533***	1.1832	2.3117**
HM alpha	0.0000	0.0575***	0.0770***	0.0542***	0.0768	0.0879***	−0.0571**	−0.0836***
HM beta	1.0000	0.7988***	0.7873***	0.8402***	0.3601***	0.3376***	0.5550***	0.4937***
HM gamma	0.0000	−0.3219***	−0.3276*	−0.2911***	−0.4634*	−0.4143***	0.1812	0.4050*

This table shows summary statistics as well as performance measures following Jensen (1968), my option-factor approach in the spirit of Coval and Shumway (2001), Broadie et al. (2009) as well as Goyal and Saretto (2009) and the time-varying beta approaches by Treynor and Mazuy (1966) as well as Henriksson and Merton (1981) based on monthly discrete returns. The first time window (Panel A) begins in February 1990 and ends in December 2003. Panel B displays the results from 2004 to 2016. The market index used in all performance regressions is the excess return of the S&P 500 total return index. Performance in terms of alpha is denoted in absolute values p.a. Estimation of standard errors is heteroscedasticity consistent according to White (1980). ***, **, * denote significance of the estimated parameter at the 1%, 5%, and 10% level, respectively.

unfairly priced options so that the realized payoffs are not congruent with the discounted expected payoffs (e.g., Constantinides et al., 2009).

Another important result to look at is the coefficient for the interaction term between the SPTR return and the crisis dummy. This estimate is mostly statistically significant in all scenarios and the sign pattern is exactly contrarious to the pattern of the crisis dummy. The interaction term denotes the additional loading of the market return in crisis periods, that is, the additional systematic risk the index is exposed to in every scenario. For the first five indices, the throughout positive coefficients indicate significantly higher market exposure during turbulent times. Significantly negative correlations for the CLL and PPUT reveal a reduction in systematic risk when scenario dummy variables take on the value one. This finding is consistent with previous results in section 4.2 where I show both analytically and empirically the dependency of systematic risk on the return of the underlying S&P 500.[15] Since crisis periods go along with stock price declines or negative returns, respectively, the sign and statistical significance of the interaction term coefficients is coherent.

4.6 | Different time periods

The first study attesting an outperformance for one of the option strategy benchmark indices is Whaley (2002) who analyzes the sample period from 1988 to 2001. The outperformance of 2.76% on an annualized basis is both statistically and economically significant. Ungar and Moran (2009) have a longer time frame available (1986–2008) and their estimate for BXM's outperformance is 1.92% p.a. It seems that the outperformance diminishes with longer windows analyzed and thus, I naively divide my sample into two equally long sub-samples. The first sample period spans from 1990 to 2003 and the second covers the time period from 2004 to the end of 2016. I repeat the calculations for the descriptive statistics and for Jensen's alpha, respectively, for each sample separately and display the results in Table 6.

At first, I compare the descriptive statistics for both time frames. An interesting outcome is that on the one hand, the average returns drop dramatically for the entire set of option indices in the second-half of the sample. On the other hand, risk in terms of volatility does not change remarkably and thus, risk-adjusted performance extremely diminishes. For example, the BXM exhibits an average return of 11.9% from 1990 to 2003, whereas from 2004 on, the mean return only amounts to 5.7%. Since the volatility changes merely by two basis points, the Sharpe Ratio falls from 0.22 to 0.13 and is considerably smaller than that of the S&P 500. It seems that positive outperformance reported in previous studies was driven by the specific sample period.

Parameter estimates of performance regression models for the early sample period are in line with previous studies on option benchmark indices. Panel A of Table 6 shows significant outperformance at the 5% level for indices that are mostly short options and peaks in BFLY's Jensen's alpha with more than 6% points on an annualized basis. In panel B of Table 6, however, there is no outperformance observable, as all alphas are close to zero and no longer statistically significant at conventional levels, except for the CLL. The initial presumption of vanishing performance in the second-half seems to be confirmed.

[15]One should note that the time-varying beta model by Henriksson and Merton (1981) is nested in scenario (v) but in addition, scenario (v) includes the dummy variable indicating negative returns.

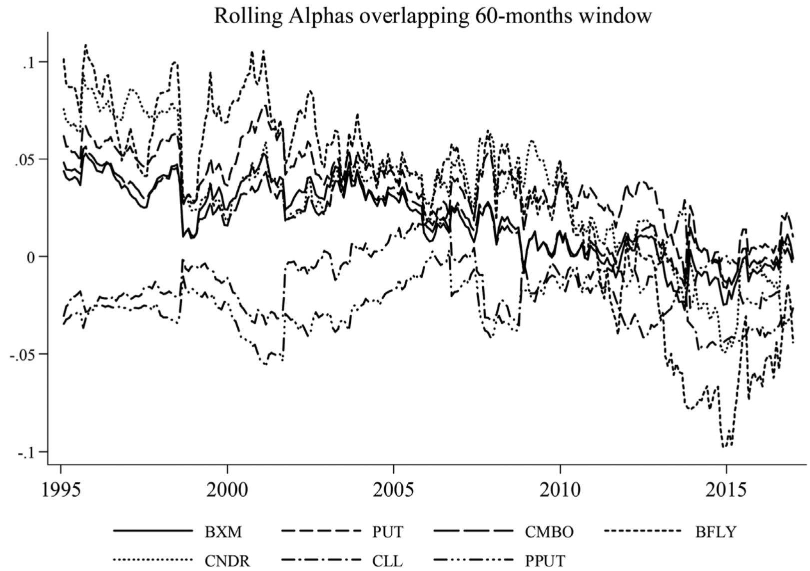

FIGURE 2 Development of rolling 60-months alphas over time. This figure shows the development of 60-months overlapping and rolling alphas for the time period from January 1995 to December 2016

Systematic risk in terms of CAPM beta also changes in the second-half of my sample, namely it increases for all indices, that is, option strategies tend to converge to the movements of the underlying S&P 500. While the increase in beta sums up to more than 13% points for the PUT (0.50–0.63), betas in the last three columns do not change significantly but fall slightly for these last three indices.

The question arising from the results obtained from Table 6 is why the results differ in such a manner. One possible explanation could be more efficient option pricing. In the first sample period, options might be richly priced and consequently, premiums earned from selling options are considered as outperformance in excess of the S&P 500. Constantinides et al. (2009) actually find evidence for overpricing of S&P 500 index options for their analyzed time horizon from 1986 to 2006. To analyze this conjecture in more detail, I estimate rolling Jensen's alphas and betas with 60-month overlapping windows. If performance diminishes over time due to, for example, more efficient option pricing, the estimated rolling alphas should reveal a dependency on time elapsed. On the other hand, performance might also be driven by varying betas or specific return regimes. Therefore, I run the following OLS regression and the results are displayed in Table 7.

$$\alpha_{\mathrm{i,t}} = \beta_{0,\mathrm{i}} + \beta_{\mathrm{SPTR,i}} r_{\mathrm{SPTR,t}} + \beta_{\beta_{\mathrm{i,t}},i}\, \beta_{\mathrm{i,t}} + \beta_{\mathrm{year,i}}\ \mathrm{year}_{\mathrm{t}} + \varepsilon_{\mathrm{i,t}} \tag{8}$$

The return of the S&P500 is negatively correlated for the first six option strategies, whereas only the BFLY reveals statistical significance. This means that, all else being equal, if the return of the underlying is negative, the performance of these strategies tends to increase. The contrary effect can be observed with the protective put index, whose estimated parameter is statistically significant and positive. The different signs are not surprising, since the PPUT and the remaining strategies are contrary.

TABLE 7 Rolling Alphas 60 months

	BXM	PUT	CLL	BFLY	CMBO	CNDR	PPUT
S&P500	−0.0241	−0.0307	−0.0028	−0.0914**	−0.0008	−0.0142	0.0446**
Beta	−0.0111	−0.0090	−0.1403***	−0.0140	−0.0215***	−0.0331***	−0.0535***
year	−0.0026***	−0.0024***	−0.0018***	−0.0069***	−0.0021***	−0.0036***	0.0007***
R^2	0.76	0.67	0.56	0.77	0.75	0.65	0.43
N	264	264	264	264	264	264	264

This table shows results for regressions of rolling alphas for all option indices on the return of the SPTR, the rolling beta as well as a year variable. The sample period spans from January 1995 to December 2016. Rolling performance and risk measures are estimated using the Jensen's (1966) approach. The dependent variable alpha is denoted on an annualized basis. Estimation of standard errors is heteroscedasticity consistent according to White (1980). ***, **, * denote significance of the estimated parameter at the 1%, 5%, and 10% level, respectively.

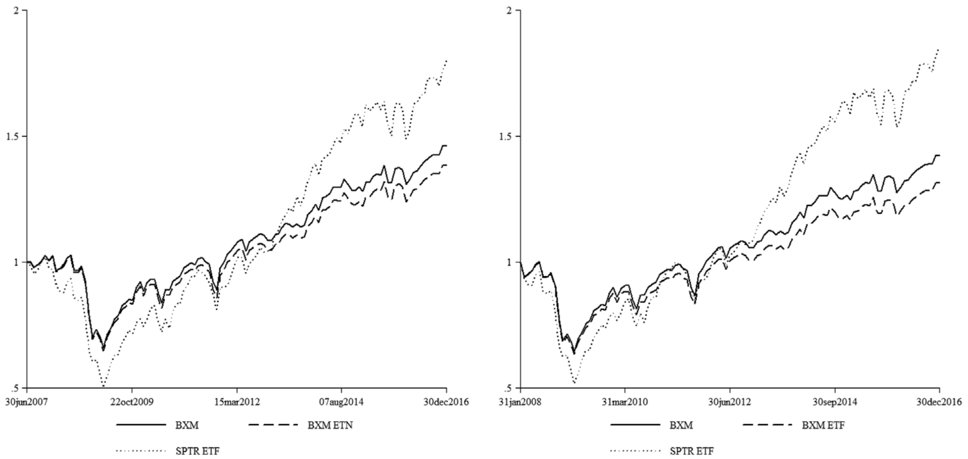

FIGURE 3 Development of $1 for different investments. This figure shows the development for $1 invested in the ipath® S&P 500 BuyWrite ETN (BXM ETN), the PowerShares S&P 500 BuyWrite ETF (BXM ETF), the BXM and the SPDR S&P 500 ETF Trust (SPDR ETF). The sample period for the BXM ETN spans from June 2007 to the end of December 2016 and the dataset for the BXM ETF starts in January 2008 ending in 2016

Loadings on the rolling 60-month beta factor are entirely negative and statistically distinguishable from zero for four indices. Alpha is, therefore, also correlated with different values of systematic risk. A very interesting outcome is the coefficient for year, which is highly significant for all option strategies. Strategies possessing mainly short option exposure, which can be derived from the loadings on my option straddle-factor in Table 4, seem to lose performance over time. In contrast to this, the sign of the coefficient for the PPUT is positive, which means that the performance increases ceteris paribus over time.

Figure 2 illustrates the development of 60-month rolling alphas over time. The performance diminishing effect is also graphically observable for most of the strategies and explains at least partly the results for the split sample analyses. The fact that alpha seems to vanish over time for option writing strategies (e.g., BXM) and underperformance for long option strategies, like the PPUT, is less pronounced is further evidence that overpricing in options as documented by Constantinides et al. (2009) diminishes with proceeding time.

5 | ANALYSIS OF INVESTABLE PRODUCTS

Although Cici and Palacios (2015) do not find any effect of options on mutual fund portfolios, Natter et al. (2016) find superior performance among U.S. domestic equity mutual funds investing in options. The outperformance is mainly driven by short option positions taken by mutual fund managers (e.g., found with covered call writing). The question is whether this phenomenon is also observable among passively managed investment vehicles that bear short option exposure.

I analyze two different investment opportunities enabling private investors to take exposure toward the oldest option writing strategy, the BXM. The products are actively traded and easily investable for private market participants at low cost. Specifically, the BXM is represented by an Exchange Traded Fund, the PowerShares S&P 500 BuyWrite ETF (BXM ETF), and an Exchange Traded Note, namely the iPath® S&P 500 BuyWrite ETN (BXM ETN).[16] Data on these products are available as early as June 2007 for the BXM ETN and January 2008 for the BXM ETF.[17]

The purpose of this analysis is to test the investments' eligibility to track the option strategy benchmark index and their performance compared to the underlying. Figure 3 shows the development of $1 invested in the respective investment vehicle or in the underlying index, respectively. Additionally, the performance of buying the S&P 500 total return index is also included via the SPDR S&P 500 ETF Trust (SPDR ETF). As the products start at different dates, the results are not directly comparable. If a private investor invests $1 in the BXM ETN in June 2007, she would end up with $1.39, whereas a hypothetical direct investment in the BXM would have yielded $1.46. A conservative and direct long position in the SPDR ETF returned a much higher dollar

[16]Exposure against the PUT can be obtained by buying WisdomTree S&P 500 PutWriteStrategy ETF (PUT ETF). However, data are only available from March 2016 onward and would result in an insufficient sample size and thus, I focus only on investments regarding the BXM.

[17]Whaley (2013) also analyzes exchange-traded products (ETP) with a focus on exposure toward the VIX.

TABLE 8 Investments' performance

	BXM ETF		BXM ETN	
	Net	**Gross**	**Net**	**Gross**
Panel A: S&P 500 TR				
Alpha	−0.0178	−0.0103	−0.0104	−0.0029
Beta	0.6694***	0.6698***	0.6784***	0.6788***
R^2	0.7987	0.7987	0.7976	0.7976
Panel B: SPDR ETF				
Alpha	−0.0175	−0.0100	−0.0100	−0.0025
Beta	0.6748***	0.6752***	0.6837***	0.6841***
R^2	0.7960	0.7960	0.7948	0.7948
Panel C: BXM				
Alpha	−0.0091***	−0.0016**	−0.0063***	0.0012*
Beta	1.0015***	1.0021***	1.0210***	1.0216***
R^2	0.9998	0.9998	0.9998	0.9998
Panel D: BXM time-varying				
Alpha TM	−0.0081***	−0.0006	−0.0079***	−0.0004
Beta TM	0.9995***	1.0001***	1.0245***	1.0251***
Gamma TM	−0.0610**	−0.0610**	0.1069***	0.1069***
R^2	0.9998	0.9998	0.9999	0.9999
Observations	108	108	115	115

This table shows performance measures following Jensen (1968) for the PowerShares S&P 500 BuyWrite ETF (BXM ETF) as well as the ipath® S&P 500 BuyWrite ETN (BXM ETN). The sample period for the BXM ETF spans from January 2008 to the end of 2016 and the dataset for the BXM ETN starts in June 2007 ending in December 2016. In Panel A, the market index used in performance regressions is the excess return of the S&P 500 total return index, in Panel B, the SPDR S&P 500 ETF Trust (SPDR ETF) serves as benchmark and in Panel C, the return of the BXM in excess of the risk-free rate is the market proxy. Panel D contains results for a time-varying beta model with the BXM as the right hand variable. Performance in terms of alpha is denoted in absolute values p.a. Estimation of standard errors is heteroscedasticity consistent according to White (1980). ***, **, * denote significance of the estimated parameter at the 1%, 5%, and 10% level, respectively.

amount of $1.80. The right graph of Figure 3 shows the dollar development for a time frame starting from January 2008 and an ETF on the BXM. The resulting final dollar amounts are $1.31 for the BXM ETF, $1.43 for the hypothetical BXM investment, and $1.86 for buying the dividend reinvesting SPDR ETF on the S&P 500. From 2007 to 2013 the S&P 500 performed worse than the option index investments since the latter have a cushion for losses in contrast to the equity index. From 2013 onward, however, the outperformance of the S&P 500 exposure is clearly superior due to capped upside potential resulting from short option positions of the strategy benchmark indices.

One dimensional performance measures reveal initial drawbacks of option strategies but do not take risk considerations into account. Therefore, I estimate Jensen's alpha in three different model approaches for both the BXM ETF and the BXM ETN, respectively. In scenario (i), the market index used is the SPTR. As fund companies demand cost for providing exchange-traded products (ETPs), in scenario (ii) the investment's return is regressed on the SPDR ETF, as it also includes costs. In scenario (iii) returns are benchmarked against the replicated BXM due to the fact that it should be the adequate benchmark.[18] For all three scenarios, I use both net and gross returns. A summary of the results can be found in Table 8.

Compared to the SPTR both the BXM ETF as well as the BXM ETN reveal an underperformance up to 1.80% points p.a. However, these measures are statistically indistinguishable from zero. Panel B displays the results for regressions with the SPDR ETF. The similarity to the figures in Panel A is not surprising since the SPDR ETF has a high correlation of 0.9977 with the SPTR and is very cheaply investable with an expense ratio of 9 basis points per year.

The BXM is by definition the adequate benchmark for these investment products and Panel C displays the outcome of the performance regressions. From the first three columns, one can see that investors' do not earn the same return as the underlying BXM, as alphas are negative and statistically significant. The gross performance of the exchange-traded note is indeed positive

[18]Using a linear performance model in this case should be unproblematic since both the investment as well as the benchmark are supposed to generate the same nonlinearities in returns.

but only hardly significant at the 10% level, and, more importantly, an investor is not able to earn the gross performance of an ETN. Although the total expense ratio is 0.75%, the ETF's performance is slightly worse than the ETN's performance.

The expense ratio is not the only driver of the ETPs' fees, as it does not include trading costs. An indication for these additional costs could be seen in the loadings on the BXM factor, which are not equal to one. These differences are statistically significant for the ETN's beta. Moreover, there are further differences between the ETF's and the ETN's betas that could arise from slightly different time periods. On the other hand, this could be due to an incongruent replication, what can also be derived from the tracking error[19] (0.18% p.a. for the ETF vs. 0.17% p.a. for the ETN).

Frino and Gallagher (2001) find that tracking error is mainly driven by market frictions such as transaction costs, that is, the tracking error might be interpreted as the cost of replication. For a deeper understanding of the replication cost of ETPs, I estimate time-varying betas for both products with the BXM as the correct benchmark, where beta is a linear function of the BXM's excess return (see Equation 9).

$$r_{\mathrm{i,t}} - r_{\mathrm{f}} = \alpha_{\mathrm{TM,i}} + \beta_{\mathrm{TM,i,t}}(r_{\mathrm{BXM,t}} - r_{\mathrm{f,t}}) + \varepsilon_{\mathrm{i,t}} \tag{9}$$

$$\text{with} \;\; \beta_{\mathrm{TM,i,t}} = b_{\mathrm{i}} + \gamma_{\mathrm{i}}(r_{\mathrm{BXM,t}} - r_{\mathrm{f,t}}) \tag{9.1}$$

The BXM is calculated using average bid and ask prices, respectively. The spreads between these two prices, however, can be of considerable magnitude as seen in the financial crisis 2008 and there is no tradable mid-price.[20] Replicating theoretically calculated indices is, therefore, non-trivial. The difficulties in copying the BXM should be observable by using time-varying betas. Suppose, the investment manager is able to follow the underlying index perfectly, then the time-varying component γ_i should be indistinguishable from zero. Contrary, a statistically significant coefficient for γ_i might indicate trading costs (e.g., transaction costs or bid/ask spreads) investment managers have to pay when replicating the underlying index.

The last panel of Table 8 displays the results of the regressions following Equation 9. The alphas for both the ETF and ETN are negative but only for net returns statistically significant. However, of more importance is the finding that the time-varying component γ_i. is statistically significant for all four columns, although the outcomes differ in signs between the ETF and the ETN. This means that the loadings on the underlying BXM index are not constant but rather vary over time and depend systematically on the return of the BXM. An interpretation of this result would be that in different time and return regimes investment managers fail to meet the exact correlation with the index due to the costs of replicating and over- or undershoot the desired return.

All in all, I conclude from this analysis that potential benefits of option-writing indices—as far as there are any—are consumed by the costs investment providers charge and have to pay for the replication, respectively. This is not coercively conflicting with the findings of Natter et al. (2016), as the outperformance of actively managed funds engaging in options the authors measure is compared to their non-using peers.

6 | CONCLUDING REMARKS

This study aims to review the benefits of option strategy benchmark indices in terms of performance and risk. I show that the performance measurement of such indices is indeed difficult and sensitive to the method used. Several problems should be addressed when analyzing the advantages and drawbacks of portfolios containing options. One finding is the time-varying delta exposure toward the underlying S&P 500 index, which might lead to severe biases if neglected. However, approaches derived from timing literature insufficiently control for this phenomenon. A solution is the methodology as proposed by Israelov and Nielsen (2015), at which the evaluator has to know the exact portfolio composition and additionally the detailed options' characteristics (delta). Analytically correct approaches as proposed by Leland (1999) and Whaley (2002) are also not able to appropriately disclose performance, as resulting performance measures do not differ remarkably from standard methods. The approach I develop in this study shows significant loadings on my option straddle-factor though it is not clear if the constants of these regressions reflect the proper performance. Option strategies are said to be protections from sharp market declines, for

[19]The tracking error is calculated following Cremers and Petajisto (2009) as the standard deviation of the residuals obtained from performance regressions using monthly net returns.

[20]I thank the anonymous referee for these helpful suggestions.

example, during times of crises and for some strategies, I am able to confirm this empirically. Another finding, which is possible due to the long sample period, is that eventual outperformance is mostly driven by the period in the first-half of my sample to the end of 2003.

Furthermore, I analyze the benefits of the BXM for private investors by measuring the performance of two directly investable products reflecting the development of this option benchmark. The conclusion, which can be drawn from this examination, is that fees charged by the investment companies and costs paid to mimic the strategy lead to an underperformance compared to the underlying option index and an incongruent replication.

ORCID

Markus Natter http://orcid.org/0000-0002-7724-0438

REFERENCES

Agarwal, V., & Naik, N. Y. (2004). Risks and portfolio decisions involving hedge funds. *Review of Financial Studies*, *17*(1), 63–98.

Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, *81*(3), 637–654.

Bookstaber, R., & Clarke, R. G. (1981). Options can alter portfolio return distributions. *Journal of Portfolio Management*, *7*(3), 63–70.

Bookstaber, R., & Clarke, R. G. (1984). Option portfolio strategies: Measurement and evaluation. *Journal of Business*, *57*(4), 469–492.

Bookstaber, R., & Clarke, R. G. (1985). Problems in evaluating the performance of portfolios with options. *Financial Analysts Journal*, *41*(1), 48–62.

Broadie, M., Chernov, M., & Johannes, M. (2009). Understanding index option returns. *Review of Financial Studies*, *22*(11), 4493–4529.

Bunnenberg, S., Rohleder, M., Scholz, H., & Wilkens, M. (2017). Jensen's alpha and the market timing puzzle. Working paper, University of Augsburg, University of Hannover, University of Nuremberg. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1253923

Chalmers, J., Kaul, A., & Phillips, B. (2013). The wisdom of crowds: Mutual fund investors' aggregate asset allocation decisions. *Journal of Banking and Finance*, *37*, 3318–3333.

Chambers, D. R., Foy, M., Liebner, J., & Lu, Q. (2014). Index option returns: Still puzzling. *Review of Financial Studies*, *27*(6), 1915–1928.

Cici, G., & Palacios, L.-F. (2015). On the use of options by mutual funds: Do they know what they are doing? *Journal of Banking and Finance*, *50*, 157–168.

Constantinides, G. M., Jackwerth, J. C., & Perrakis, S. (2009). Mispricing of S&P 500 index options. *Review of Financial Studies*, *22*(3), 1247–1277.

Coval, J. D., & Shumway, T. (2001). Expected option returns. *Journal of Finance*, *56*(3), 983–1009.

Cremers, K. J. M., & Petajisto, A. (2009). How active is your fund manager? A new measure that predicts performance. *Review of Financial Studies*, *22*(9), 3329–3365.

Feldman, B. E., & Roy, D (2005). Passive options-based investment strategies—The case of CBOE S&P 500 buy write index. *Journal of Investing*, *14*(2), 66–83.

Ferson, W. E., & Schadt, R. W. (1996). Measuring fund strategy and performance in changing economic conditions. *Journal of Finance*, *51*(2), 425–461.

Frino, A., & Gallagher, D. R. (2001). Tracking S&P 500 index funds. *Journal of Portfolio Management*, *28*(1), 44–55.

Fung, W., & Hsieh, D. A. (2004). Hedge fund benchmarks: A risk-based approach. *Financial Analysts Journal*, *60*(5), 65–80.

Goetzmann, W., Ingersoll, J., Spiegel, M., & Welch, I. (2007). Portfolio performance manipulation and manipulation-proof performance measures. *Review of Financial Studies*, *20*(5), 1503–1546.

Goyal, A., & Saretto, A. (2009). Cross-section of option returns and volatility. *Journal of Financial Economics*, *94*(2), 310–326.

Henriksson, R. D., & Merton, R. C. (1981). On market timing and investment performance. II. Statistical procedures for evaluating forecasting skills. *Journal of Business*, *54*(4), 513–533.

Israelov, R., & Klein, M. (2016). Risk and return of equity index collar strategies. *Journal of Alternative Investments*, *19*(1), 41–54.

Israelov, R., & Nielsen, L. N. (2015). Covered calls uncovered. *Financial Analysts Journal*, *71*(6), 44–57.

Jagannathan, R., & Korajczyk, R. A. (1986). Assessing the market timing performance of managed portfolios. *Journal of Business*, *59*(2), 217–235.

Jagannathan, R., & Wang, Z. (1996). The conditional CAPM and the cross-section of expected returns. *Journal of Finance*, *51*(1), 3–53.

Jensen, M. C. (1968). The performance of mutual funds in the period 1945–1964. *Journal of Finance*, *23*(2), 389–416.

Kapadia, N., & Szado, E. (2007). The risk and return characteristics of the buy-write strategy on the Russell 2000 Index. *Journal of Alternative Investments*, *9*(4), 39–56.

Leland, H. E. (1999). Beyond mean-variance: Performance measurement in a nonsymmetrical world. *Financial Analysts Journal*, *55*(1), 27–36.

Lhabitant, F. S. (2000). Derivatives in portfolio management: Why beating the market is easy. *Derivatives Quaterly*, *7*(2), 39–45.

Merton, R. C., Scholes, M., & Gladstein, M. (1978). The returns and risk of alternative call option portfolio investment strategies. *Journal of Business*, *51*(2), 183–242.

Merton, R. C., Scholes, M., & Gladstein, M. (1982). The returns and risks of alternative put-option portfolio investment strategies. *Journal of Business*, *55*(1), 1–55.

Natter, M., Rohleder, M., Schulte, D., & Wilkens, M. (2016). The benefits of option use by mutual funds. *Journal of Financial Intermediation*, *26*, 142–168.

Pounds, H. M. (1978). Covered call option writing: Strategies and results. *Journal of Portfolio Management*, *4*(2), 31–42.

Schulte, D., & Stamos, M. Z. (2015). The performance of equity index option strategy returns during the financial crisis. Working paper, Tecta Invest GmbH, Allianz Deutschland.

Sharpe, W. F. (1966). Mutual fund performance. *Journal of Business*, *39*(1), 119–138.

Sortino, F. A., & Price, L. N. (1994). Performance measurement in a downside risk framework. *Journal of Investing*, *3*(3), 59–64.

Stutzer, M. (2000). A portfolio performance index. *Financial Analysts Journal*, *56*(3), 52–61.

Treynor, J., & Black, F. (1972). Portfolio selection using special information, under the assumptions of the diagonal model, with mean-variance portfolio objectives, and without constraints. In G. P. Szego & K. Shell (Eds.), *Mathematical methods in investment and finance 4* (pp. 367–384). Amsterdam: North-Holland.

Treynor, J., & Mazuy, K. (1966). Can mutual funds outguess the market? *Harvard Business Review*, *44*(4), 131–136.

Ungar, J., & Moran, M. (2009). The cash-secured putwrite strategy and performance of related benchmark indexes. *Journal of Alternative Investments*, *11*(4), 43–56.

Whaley, R. E. (2002). Return and risk of CBOE buy write monthly index. *Journal of Derivatives*, *10*(2), 35–42.

Whaley, R. E. (2013). Trading volatility: At what cost? *Journal of Portfolio Management*, *40*(1), 95–108.

White, H. (1980). A Heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroscedasticity. *Econometrica*, *48*(4), 817–838.

How to cite this article: Natter M. Options-based benchmark indices—A review of performance and (in)appropriate measures. *J Futures Markets*. 2018;38:271–288. https://doi.org/10.1002/fut.21865