Econometrics & Management Science - Quantitative Finance

Erasmus University Rotterdam

Master Thesis

Modeling the IVS of Leveraged ETF Options Based on the Data of Information-Rich and Highly Liquid ETF Options.

Author:
Sjoerd Sterk (431754)

Supervised by:
Xun (X.) Gong (Erasmus University)
Xiao (X.) Xiao (Erasmus University)

August 16, 2018

ABSTRACT

In this research we attempt to model and forecast the implied volatility surface of LETF options by utilizing the information-rich and highly liquid ETF option data. The main focus here lies on assessing the performance of the method proposed by Leung & Sircar (2015), as their own empirical research is limited. Furthermore we adopt the method proposed by Bernales & Guidolin (2014) on the combination of LETF and ETF options, as this has not been done before in existing literature. These two models and several benchmark models are tested both in a statistical and economical setting. Overall this research shows that the Leung & Sircar method works in practice, however we were unable to find any general occasions in which the Leung & Sircar method should be preferred over the use of the direct estimation approach. The adoption of the Bernales & Guidolin method on the combination of ETF and LETF options has proven more useful, as it outperforms the direct estimation approach is several cases, both in statistical and economical sense. However, we have to note that the results are rather close and therefore we can not make any definite conclusions on this matter.

Keywords: Implied Volatility Surface (IVS), Leveraged Exchange Traded Fund (LETF) Options, Trading Strategies, Predictability, Moneyness Scaling.

TABLE OF CONTENTS

1. INTRODUCTION

Exchange-traded funds (ETFs) are products that replicate the returns on various financial quantities. In 1993 the first ETFs became available in the United States and since 1999 European investors are likewise able to benefit from this new investment vehicle. The main benefit of an ETF is that it combines the valuation feature of a mutual fund with the tradability feature of a closed-end fund. The number of available ETFs worldwide has grown from 276 in 2003 to 4779 by the end of 2016, nowadays covering almost every market sector, niche and trading strategy. Another variety of ETFs are the leveraged ETFs (LETFs) that, as the name suggests, promise a fixed leverage ratio on its returns. For instance the ProShares Ultra S&P500 (SSO) has a β of $+2$, generating twice the daily returns on the S&P500 minus an expense fee. LETFs have not been around for long, as they were not introduced until the summer of 2006. Even though the financial crisis has given rise to a large debate of deleveraging, the market for LETFs has been growing. Despite this growth the available literature on this matter is still limited. Cheng & Madhavan (2010) provide a framework that allows for a better understanding of the underlying dynamics of leveraged and inverse (leveraged) ETFs. They show that the daily re-leveraging of LETFs causes extra volatility on the markets. Avellaneda & Zhang (2010) further researched the path-dependence of LETF returns and conclude that leveraged funds can be used to replicate the returns of the underlying index, provided that a dynamic rebalancing strategy is used. This is based on the idea that ETFs and their corresponding LETFs are based on the same underlying and therefore they also share the same source of randomness. This should, in theory, allow for consistent pricing.

The availability of options on LETFs even allows for an additional layer of leveraging. Due to the associated increased volatility of leveraging it is of great importance for investors to fully comprehend the risks of such investment vehicles. In the literature there are two popular option pricing models, namely the Black & Scholes (1973) pricing model (BS model) for European options and the binomial tree model, as proposed by Cox et al. (1979), for American option. We can numerically invert the BS model and use the option price as stated by the market to calculate the so-called implied volatility (IV). In practice option prices are often converted into IVs, because prices of different option contracts with the same underlying are difficult to compare. Contrary to the early assumption of Black & Scholes (1973) the volatility of different option contracts written on the same underlying asset is not constant. It turns out that the IV from option contracts written on the same underlying asset differ across two different dimensions; the strike prices and the time-to-maturities. This anomaly is known as the implied volatility surface (IVS). We note is that throughout this report the abbreviation IVs refers to implied volatilities, whereas IVS refers to implied volatility surface. In the dimension of the strike price the shape is presented by the famous volatility smile, whereas the the shape in the time-to-maturity dimension is caused by the volatility term structure. A lot of research has been done on modeling the dynamics of the IVS, e.g. by Dumas et al. (1998), Cont et al. (2002), Goncalves & Guidolin (2006) and Fengler et al. (2007). Most of the papers written on the subject of the IVS conducted their research on index options, since these options have the highest liquidity. Equity options tend to have a much lower liquidity and therefore it can be much harder to model the dynamics of its IVS. However, Bernales & Guidolin (2014) present evidence of dynamic linkages between the IVS of equity and index options. They built an approach of utilizing the combination of index option data and equity option data to model the dynamics of the IVS of equity options. Despite its importance to investors, the current availability of research on the IVS of LETFs is still very limited. Leung & Sircar (2015) pioneered in this area by introducing the concept of moneyness scaling. This new formula approximately links the option implied volatilities between leveraged and unleveraged ETFs by transforming the moneyness coordinates. Furthermore, they apply asymptotic techniques under a multiscale stochastic volatility framework to derive an approximation for both the LETF option price and its IV. They implement this theory in practice by first calibrating the stochastic volatility group parameters using unleveraged ETF option data and then inferring the approximated IV of LETF options. Their study lacks further research on the dynamics of these stochastic volatility group parameters. Härdle et al. (2016) studied the statistical properties of the moneyness scaling transformation by constructing bootstrap uniform confidence bands. They also presented a more sophisticated approach for modeling the IVS of LETF options through the use of a dynamic semi-parametric model (DSFM). By utilizing a VAR model to capture the dynamics of the stochastic factor loadings they allow for out-of-sample forecasting. They exploit information from these forecasts by using a statistical arbitrage strategy on empirical data, however their dataset only spans 55 days.

The contribution of this master thesis is to model the IVS of LETF options by utilizing the data of the highly liquid and information-rich unleveraged ETF options. We work towards this research goal in several steps. First we search for a way to allow for comparison between the IV data of LETF options and the regular ETF options. Next we assess the usefulness of the Leung & Sircar method when modeling and forecasting the IVS of LETF options based on the information of the IVS of ETF options? Furthermore we adopt the Bernales & Guidolin method on the combination of LETF and ETF options to model and forecast the IVS of said LETF options. Lastly we search for a statistical arbitrage strategy to exploit the pricing discrepancies found by modeling the IVS of LETF options based on the ETF option data. It has to be noted that the mentioned pricing discrepancies can be found in the discrepancies of the IVs of options, as in practice it is common to refer to the IV when talking about the price of an option.

First of all, we use the moneyness scaling approach of Leung & Sircar (2015) to allow for comparison between the IV data of ETF and LETF options. Secondly, we adopt the Leung & Sircar method by estimating a deterministic model on ETF option data, using these estimates to calibrate stochastic volatility parameters and utilizing these calibrated parameters to find the estimates of the deterministic model for LETF options as discussed in Leung & Sircar (2015). We extend their approach by modeling the dynamics of the estimates deterministic model for LETF options using a VAR model. Thirdly, as a benchmark we adopt the method of Bernales & Guidolin, whom presented an approach of utilizing the combination of index option data and equity option data to model the dynamics of equity options. In a similar fashion it should be possible to use data of unleveraged ETF options to contribute to a more accurate model for the IVS of options written on LETFs. This is based on the idea that the data of unleveraged ETF options tends to be significantly richer than the associated LETF option data. To allow for comparison between the aforementioned models we utilize the same deterministic model as used for the Leung & Sircar method. Furthermore we use three other benchmark models. Among these benchmark models is a direct estimation approach, which estimates the coefficients of our deterministic model directly on the LETF IVS data. All models are assessed on their out-of-sample forecasting performance using several performance measures. Besides this statistical evaluation we analyze the used models in the economical sense by testing whether it is possible to exploit the information provided by the models by using a trading strategy. For this so called statistical arbitrage strategy we aim to built a strategy based on what we expect will happen; its forecasted values.

To conduct our research we use options on the very liquid and information-rich ETF that tracks the S&P500 index (SPY) and the options on its corresponding LETFs, for which we use ProShares Ultra S&P500 (SSO), ProShares UltraShort S&P500 (SDS), ProShares UltraPro S&P500 (UPRO) and ProShares UltraPro Short S&P500 (SPXU), with betas of +2, -2, +3 and -3 respectively. Our dataset stretches from July 6th 2009 to April 29th 2016. Overall our research shows that the Leung & Sircar method works in practice, however we were unable to find any general occasions in which the Leung & Sircar method should be preferred over the use of the direct estimation approach. Perhaps the method proofs more useful when adopted on a dataset with extremely illiquid LETF options, because in this case the direct estimation approach will not be able to fully capture the dynamics of the IVS of LETF options due to lack of data points. The adoption of the Bernales & Guidolin method on the combination of ETF and LETF options has proven more useful, however its results are very close to the results of the direct estimation approach. Therefore, we suggest that more empirical research should be conducted on the usability of this method on the combination of LETF and ETF options.

This research aims to provide further insights on modeling and forecasting the IVS of LETF options based on the information-rich and highly liquid ETF options. The main focus lies on assessing the performance of the Leung & Sircar method, as their own empirical research is very limited. For instance our research is conducted on a dataset of nearly 7 years, much larger than the aforementioned study on the IV of LETF options. Furthermore they did not extend their method towards a dynamic setting, in which the model can be used to provide forecasts of the IVS. Another important contribution of our research is the adoption of the Bernales & Guidolin method on the combination of LETF and ETF options, as this has not been done before in existing literature. Thus, our research contributes to the current limited availability of literature on the IVS of options written on LETFs by both comparing and extending the existing literature.

The rest of this thesis is organized as follows. Section 2 describes the methods that are used for our research in detail. Section 3 describes the data that we conduct our research on and several preliminary results. Section 4 presents the results of our empirical analysis. Finally section 5 concludes and section 6 provides several recommendations.

2. METHODS

In this section we describe the methods that are used to conduct our research. To be able to use the data on regular ETF options we have to allow for comparison between the IV of LETF options and their regular counterpart. Therefore, we first discuss the necessary methods to allow for this comparison. In the next part of our research we describe various models that aim to accurately describe the IVS of LETF options based on the information of the regular ETF options. Furthermore we lay-out an evaluation framework based on several evaluation methods to test the models both in statistical and economical sense.

2.1 Comparison of (L)ETF options

2.1.1 Implied Volatility Scaling

To allow for comparison between the IV of LETF options and their regular counterpart we have to ensure that the implied volatilities are on the same scale by multiplying the IV of the LETF by the inverse of its corresponding β. Since volatilities cannot be negative we use the absolute value of the β. This scaling procedure is common in literature on LETF options, e.g. both Leung & Sircar (2015) and Härdle et al. (2016) use it. The idea behind it can be seen in appendix A. The transformed IV, which is on the same scale as the IV of regular ETF options, has the following representation:

$$\widetilde{IV}_{\text{LETF}} = \frac{1}{|\beta_{\text{LETF}}|} IV_{\text{LETF}} \tag{1}$$

In the rest of the report, when talking about the IV of LETF options, we refer to this scaled variant.

2.1.2 Moneyness Scaling

Leung & Sircar (2015) were the first to propose a moneyness scaling method to allow for an approximate comparison between options on ETFs and LETFs. A summary of their derivations can be found in appendix B. For the moneyness measure they use the log-moneyness, which is defined by: $LM = \log(K/L_t)$, where K is the strike price and L_t is the price of the underlying at time t. For two LETFs, LETF_1 and LETF_2, with different leverage ratios, β_1 and β_2, the formula that links them on the moneyness scale of LETF_1 takes the following approximate form:

$$\widetilde{LM}_{\text{LETF}_2} = \frac{\beta_1}{\beta_2}\left[LM_{\text{LETF}_2} + (r(\beta_2 - 1) + c_2)\tau + \frac{\beta_2(\beta_2 - 1)}{2}\bar{\sigma}^2\tau\right] - (r(\beta_1 - 1) + c_1)\tau - \frac{\beta_1(\beta_1 - 1)}{2}\bar{\sigma}^2\tau \tag{2}$$

where r is the risk-free rate over the duration of the option, c_i is the associated yearly fee with LETF_i, τ is the time-to-maturity and $\bar{\sigma}$ is the average of the implied volatilities of the ETF option. A regular ETF can be thought of as a LETF with $\beta_1 = 1$. Furthermore, it has to be noted that the fee for a regular ETF is much smaller than the fee for its corresponding LETF, thus near zero and therefore we assume $c_1 = 0$. This reduces equation 2 to:

$$\widetilde{LM}_{\text{LETF}} = \frac{1}{\beta_2}\left[LM_{\text{LETF}} + (r(\beta_2 - 1) + c_2)\tau + \frac{\beta_2(\beta_2 - 1)}{2}\bar{\sigma}^2\tau\right] \tag{3}$$

where $\widetilde{LM}_{\text{LETF}}$ is on the same moneyness scale as the regular ETF, β_2 is the leverage ratio of the LETF and c_2 its associated yearly fee.

2.2 Modeling the IVS of LETF Options

In order to accurately describe the IVS of LETF options based on the information of the regular ETF options we have to model the IVS and its dynamics. The used models are tested both in a statistical and economical sense by forecasting the IVS using a moving-window approach. For the statistical performance we compare their forecast performance on different measures. For the economical sense the models are tested by checking their performance on a trading strategy that is based on statistical arbitrage. First of all we start off by describing the models that are used throughout this research.

2.2.1 Leung & Sircar Method (LS)

Besides the affine moneyness-scaling formulas, Leung & Sircar (2015) describe a more sophisticated way to link the IVS of LETF options to the IVS of ETF options. Under the multiscale stochastic volatility framework they work out a derivation of the first order approximation for the IV of a regular ETF option. The class of option pricing models that is used for their analysis comes from Fouque et al. (2011), and thus further technical details and derivations can be found there. For those that are interested, we present a summary of these derivations in appendix C. These derivations are a bit tedious, but please note that they do not have to be completely understood to be able to use the Leung & Sircar method in practice.

Following from their derivations we present the first order approximation for the IV of ETF options, which at one point in time is presented by:

$$IV_{\text{ETF}} = b_1 + b_2\tau + a_1\frac{LM_{\text{ETF}}}{\tau} + a_2 LM_{\text{ETF}} + \epsilon \tag{4}$$

where the parameters (b_1, b_2, a_1, a_2) are defined in terms of the stochastic volatility group parameters (σ, V_0, V_1, V_3) by the following formulas:

$$\begin{aligned} b_1 &= \sigma + \frac{V_3}{2\sigma}\left(1 - \frac{2r}{\sigma^2}\right) & a_1 &= \frac{V_3}{\sigma^3} \\ b_2 &= V_0 + \frac{V_1}{2}\left(1 - \frac{2r}{\sigma^2}\right) & a_2 &= \frac{V_1}{\sigma^2} \end{aligned} \tag{5}$$

where (a_1, a_2) are the the skew slope parameters and (b_1, b_2) the level parameters. Furthermore r is the yearly risk-free rate.

They also derived the first order approximation for the IV of LETF options, which at one point in time is given by:

$$\widetilde{IV}_{\text{LETF}} = b_{1,\beta} + b_{2,\beta}\tau + a_{1,\beta}\frac{LM_{\text{LETF}}}{\tau} + a_{2,\beta} LM_{\text{LETF}} + \epsilon \tag{6}$$

Note that the log-moneyness in this approximation is the unscaled variant. The parameters $(b_{1,\beta}, b_{2,\beta}, a_{1,\beta}, a_{2,\beta})$ are defined in terms of the same stochastic volatility group parameters (σ, V_0, V_1, V_3) as its unleveraged variant. Leung & Sircar (2015) derived the following formulas to show how the β interacts with the stochastic volatility group parameters:

$$\begin{aligned} b_{1,\beta} &= \sigma + \frac{\beta V_3}{2\sigma}\left(1 - \frac{2r}{\beta^2\sigma^2}\right) & a_{1,\beta} &= \frac{V_3}{\beta\sigma^3} \\ b_{2,\beta} &= V_0 + \frac{\beta V_1}{2}\left(1 - \frac{2r}{\beta^2\sigma^2}\right) & a_{2,\beta} &= \frac{V_1}{\beta\sigma^2} \end{aligned} \tag{7}$$

Leung & Sircar (2015) estimate (b_1, b_2, a_1, a_2), using OLS on the information rich and highly liquid ETF option data by means of equation 4. However, as suggested by Hentschel (2003) it is preferable to use General Least Squares (GLS) instead of OLS to estimate the parameters in this linear model. The reason for this is the fact that there are pervasive measurement errors in implied volatilities that may introduce heteroskedasticity and autocorrelation in the OLS residuals. Hence we use iterative feasible GLS to estimate these parameters. Using these estimates, $(\hat{b}_1, \hat{b}_2, \hat{a}_1, \hat{a}_2)$, it is possible to calibrate the stochastic volatility group parameters $(\sigma, V_0, V_1), V_3)$. This can be done by rewriting the formulas of equation 5 towards:

$$\begin{aligned} \hat{\sigma} &= \frac{\hat{a}_1 r + \hat{b}_1}{1 + \frac{\hat{a}_1}{2}} & \hat{V}_3 &= \hat{a}_1\sigma^3 \\ \hat{V}_1 &= \hat{a}_2\sigma^2 & \hat{V}_0 &= \hat{b}_2 - \frac{V_1}{2} + \frac{rV_1}{\sigma^2} \end{aligned} \tag{8}$$

Next these calibrated parameters can be used to find the estimates $(\hat{b}_{1,\beta}, \hat{b}_{2,\beta}, \hat{a}_{1,\beta}, \hat{a}_{2,\beta})$. To capture the dynamics of the IVS we use the above procedure for each point in time, which provides us with the estimates $\hat{\gamma}_{\beta,t} = (\hat{b}_{1,\beta,t}, \hat{b}_{2,\beta,t}, \hat{a}_{1,\beta,t}, \hat{a}_{2,\beta,t})'$ for $t = 1, ..., T$. The dynamics of these estimates are modeled by using a VAR(p) model, which is given by:

$$\hat{\gamma}_{\beta,t+1} = \alpha + \sum_{j=0}^{p} \Phi_j \hat{\gamma}_{\beta,t-j} + u_t \tag{9}$$

In which the number of lags (p) is chosen by minimizing the Bayesian Information Criteria (BIC) with a preselected maximum of three lags. This provides us with a way to forecast these estimates, which in their turn can be used to forecast the scaled IV of the LETF options by means of the following equation:

$$\widehat{\widetilde{IV}}_{\text{LETF},t+1} = \hat{b}_{1,\beta,t+1} + \hat{b}_{2,\beta,t+1}\tau + \hat{a}_{1,\beta,t+1}\frac{LM_{\text{LETF}}}{\tau} + \hat{a}_{2,\beta,t+1}LM_{\text{LETF}} \tag{10}$$

2.2.2 Benchmark Models

To be able to compare the forecasting performance of the Leung & Sircar (LS) method we came up with several benchmark models. The third benchmark model is a simple random walk approach directly on the IV data. The other three benchmark models provide different methods to model the dynamics of the estimates, which in their turn can be forecasted and used in equation 10 to provide the IV forecasts. For this research we use the following benchmark models:

1. **Direct Estimation (DE):** To test whether the use of calibrating the volatility group parameters based on the ETF options improves the model performance we directly estimate the first order approximation directly on the LETF option data by means of the same deterministic function as used in equation 6 by the Leung & Sircar method. In this case our regression has the following representation:

$$\widetilde{IV}_{\text{LETF}} = \beta_0 + \beta_1\tau + \beta_2\frac{LM_{\text{LETF}}}{\tau} + \beta_3 LM_{\text{LETF}} + \epsilon \tag{11}$$

   Likewise, the estimation is done by means of feasible GLS, which provides us with the following estimates: $\hat{\gamma}_{DE,t} = (\hat{\beta}_{0,t}, \hat{\beta}_{1,t}, \hat{\beta}_{2,t}, \hat{\beta}_{3,t})'$ for $t = 1, ..., T$. In the same way as for the Leung & Sircar method the dynamics of the IVS are captured by using a VAR(p) model on $\hat{\gamma}_{DE,t}$. Equivalently we minimize the BIC to select the number of lags (p).

$$\hat{\gamma}_{DE,t+1} = \alpha + \sum_{j=0}^{p} \Phi_j \hat{\gamma}_{DE,t-j} + u_t \tag{12}$$

2. **Bernales & Guidolin (BG):** Bernales & Guidolin (2014) present evidence of dynamic linkages between the IVS of equity and index options. Their idea is based on the fact that index option data tends to be significantly richer than the equity option data. This is caused by the higher liquidity of index options as well as the availability of more different option contracts (differing on both strike price and time-to-maturity). They use a two-step approach to model the dynamics of the IVS. First they model the IVS of both equity options and index options seperately using a deterministic linear model. Next, they use a simple VARX structure to model the dynamics of the estimated coefficients of the equity options by including the estimated coefficients of the index options as exogenous variables. This way they produce accurate predictions that both statistically and economically outperform predictions produced by competing models of common use in literature. However, it has to be noted that the trading profits disappear when transaction costs are taken into account. Analogously, their simple approach can be adopted for LETF options. To ensure comparability we use the same deterministic function as used in equation 6 by the Leung & Sircar method. First we estimate the deterministic linear function for the ETF options by means of GLS on:

$$\widetilde{IV}_{\text{ETF}} = \alpha_0 + \alpha_1\tau + \alpha_2\frac{LM_{\text{LETF}}}{\tau} + \alpha_3 LM_{\text{LETF}} + \epsilon \tag{13}$$

This provides us with the following estimates: $\hat{\delta}_{BG,t} = (\hat{\alpha}_{0,t}, \hat{\alpha}_{1,t}, \hat{\alpha}_{2,t}, \hat{\alpha}_{3,t})'$. Next we estimate the deterministic linear function for the LETF options by means of GLS in the same way as the direct estimation method on equation 11. This provides us with the following estimates: $\hat{\gamma}_{BG,t} = (\hat{\beta}_{0,t}, \hat{\beta}_{1,t}, \hat{\beta}_{2,t}, \hat{\beta}_{3,t})'$. Now the dynamics of these estimates are modeled by a VARX(p,q) model, where the estimates $\hat{\delta}_{BG,t}$ are adopted as exogenous variables. We select the number of lags (p and q) by minimizing the BIC.

$$\hat{\gamma}_{BG,t+1} = \alpha + \sum_{j=0}^{p} \Phi_j \hat{\gamma}_{BG,t-j} + \sum_{k=1}^{q} \Psi_k \hat{\delta}_{BG,t-k} + u_t \tag{14}$$

3. **Random walk (RW):** To test to which degree our models are able to capture the dynamics of the IV of LETF options we use a random walk approach directly on the IV data, which is represented by:

$$\widehat{\widetilde{IV}}_{\text{LETF},t+1} = \widetilde{IV}_{\text{LETF},t} \tag{15}$$

4. **Strawman (SM):** We use an ad-hoc 'Strawman' model, which is a random walk approach on the estimates of the Leung & Sircar method. This allows us to test whether the VAR structure proves useful to capture the dynamics of the estimates.

$$\hat{\gamma}_{SM,t+1} = \hat{\gamma}_{\beta,t} \tag{16}$$

2.3 Statistical Model Evaluation

The used models are tested in statistical sense by testing their in-sample fit and the performance of out-of-sample forecasts. The latter is done using recursively by using a moving-window approach. The true IV is the scaled IV of the LETF options, previously denoted as $\widetilde{IV}_{\text{LETF},t+1}$. For this section we define $Y_{t+1,j}$ as the true IV and $\hat{Y}_{t+1,j|t}$ as the forecasted value, where $j = 1, ..., J_{t+1}$ is the option numbering on day $t + 1$. We compare the forecast performance on three different criteria. These are calculated for each time point in the evaluation period. For each model we take the average of each measure over all time points.

1. **Root Mean Squared Prediction Error (RMSPE):** We take the square root of the averaged squared deviation between the true IV and the one-step ahead IV forecast conditional on time t. It has the following representation:

$$\text{RMSPE} \stackrel{\text{def}}{=} \sqrt{\frac{1}{J_{t+1}} \sum_{j=1}^{J_{t+1}} (Y_{t+1,j} - \hat{Y}_{t+1,j|t})^2} \tag{17}$$

2. **Mean Absolute Prediction Error (MAPE):** Kind of like the RMSPE, but this time we take the absolute value of the deviation. It has the following representation:

$$\text{MAPE} \stackrel{\text{def}}{=} \frac{1}{J_{t+1}} \sum_{j=1}^{J_{t+1}} \left| Y_{t+1,j} - \hat{Y}_{t+1,j|t} \right| \tag{18}$$

3. **Mean Correct Prediction of Direction of Change (MCPDC):** This measure provides an insight into how often the model predicted the direction of change, expressed as an percentage of all forecasts. It has the following representation:

$$\text{MCPDC} \stackrel{\text{def}}{=} \frac{1}{J_{t+1}} \sum_{j=1}^{J_{t+1}} I_{sign(Y_{t+1,j} - Y_{t,j}) = sign(\hat{Y}_{t+1,j|t} - Y_{t,j})} \tag{19}$$

Here I is an indicator function that has value 1 when its argument is true and 0 otherwise. The argument checks whether the model correctly forecasted the direction of change for the next day. Both the moneyness and time to maturity of an option change over the course of each day. As we compare $t + 1$ and t it is important to note that we assume this change to be negligible.

We use both the RMSPE and MAPE, because they act differently on possible outliers. The aforementioned performance statistics are also reported for different segments of the IVS. This is done because it is likely that the forecasting performance of each individual model varies across different regions of the IVS.

2.4 Economical Model Evaluation

In this section we aim to evaluate the models in economical sense by back-testing a statistical arbitrage strategy on our dataset. It's evident that the aforementioned MCPDC statistic is of great importance for this matter, as it is highly correlated to the profits of any statistical arbitrage strategy. We built our statistical arbitrage strategy on the assumption that the IV of LETF options will reverse towards its forecasted value, whereas in our case we use the forecasted IVS to construct a delta-hedged option portfolio. Such a strategy aims to exploit the information from the discrepancies between the true LETF IVS and its forecasted counterpart. This is done by comparing the forecasted IV, denoted by $\widehat{IV}_{t+1|t}$, with the true scaled IV of the LETF, denoted by $\widetilde{IV}_t$.

We adopt the same strategy, with some slight adjustments, as used by Härdle et al. (2016). We start off by finding the unique time-to-maturities that exist for the LETF options. Next we select the corresponding string of data for one of these time-to-maturities. For this string of data we calculate the difference between the forecasted IVs and their corresponding true scaled IVs ($\widehat{IV}_{t+1|t} - \widetilde{IV}_t$). For each string of data there are three different cases that can occur. For all three cases we use the underlying LETF asset to make the whole portfolio delta-neutral, as this protects the portfolio against unfavorable moves in the underlying asset. We terminate all of our portfolios at time $t+1$. The way we handle these three cases, their corresponding net construction costs and their one-day net gains are presented below. In all cases $C_{t,+}$ is the price at time t of the option contract that we are taking a long position on. Its corresponding delta is denoted by $\Delta_{C_{t,+}}$. In the same way $C_{t,-}$ is the price of the option contract that we are taking a short position on and $\Delta_{C_{t,-}}$ is used for its corresponding delta. Lastly S_t is the price of the underlying contract at time t.

1. If $\widehat{IV}_{t+1|t} > IV_t$ for the whole string, then long the option contract corresponding to the largest absolute positive difference (as we expect the IV of this particular contract to increase the most). In this case the net construction cost of the portfolio at time t is:
$$V_t = C_{t,+} - \Delta_{C_{t,+}} S_t$$
This results in the following net one-day gain:
$$G_{t+1} = (C_{t+1,+} - C_{t,+}) - \Delta_{C_{t,+}}(S_{t+1} - S_t)$$

2. If $\widehat{IV}_{t+1|t} < IV_t$ for the whole string, then short the option contract corresponding to the largest absolute negative difference (as we expect the IV of this particular contract to decrease the most). In this case the net construction cost of the portfolio at time t is:
$$V_t = -(C_{t,-} - \Delta_{C_{t,-}} S_t)$$
This results in the following net one-day gain:
$$G_{t+1} = -(C_{t+1,-} - C_{t,-}) + \Delta_{C_{t,-}}(S_{t+1} - S_t)$$

3. If $\widehat{IV}_{t+1|t}$ and IV_t intersect, then long the option contract with the largest absolute positive difference (as we expect the IV of this particular contract to increase the most) and short the option contract with the largest absolute negative difference (as we expect the IV of this particular contract to decrease the most). In this case the net construction cost of the portfolio at time t is:
$$V_t = (C_{t,+} - C_{t,-}) - (\Delta_{C_{t,+}} - \Delta_{C_{t,-}})S_t$$
Notice that we short the underlying asset if the delta of the option that we long is larger than the delta of the option that we short ($\Delta_{C_{t,+}} > \Delta_{C_{t,-}}$), otherwise we long the underlying asset. This results in the following net one-day gain:
$$G_{t+1} = (C_{t+1,+} - C_{t,+}) - (C_{t+1,-} - C_{t,-}) - (\Delta_{C_{t,+}} - \Delta_{C_{t,-}})(S_{t+1} - S_t)$$

Härdle et al. (2016) restrict their strategy to be able to buy a maximum of 2 option contracts per day, which means they long 2 option contracts in the first case, short 2 contracts in the second case and both short and long a contract in the third case. However, this method puts more weight on contracts that have a relatively high price when investigating the aggregated trading profits. Therefore we choose to invest a fixed amount of money per day, which is in line with the economical evaluation as performed by Bernales & Guidolin (2014). Thus, we start off with initial capital C_0 and use all of it to construct our first portfolio. In the scenario that the net cost of the portfolio is positive ($V_t > 0$) we purchase the quantity $X_t = C_0/V_t$ in units of the delta-hedged portfolio. In this case our total one-day net gain is: $TG_{t+1} = X_t \cdot G_{t+1}$. However, when the net cost of the portfolio is negative ($V_t < 0$) we sell the quantity $X_t = C_0/|V_t|$ in units of the delta-hedged portfolio, which generates a cash inflow of C_0. We invest this C_0 together with our initial C_0 at the riskless interest rate over one day. In this case our total one-day net gain is: $TG_{t+1} = 2C_0 \cdot (e^{(r_t/252)} - 1) - X_t \cdot G_{t+1}$. We set up our next portfolio using only our initial capital C_0. Thus, in case our total one-day net gain is positive we set this profit aside and when the total one-day net gain is negative we replenish our capital back to its initial value from an external fund. This process is repeated until time T, after which we can calculate the total profit/loss of our trading strategy.

3. DATA AND PRELIMINARY RESULTS

This section gives a general overview of the dataset that is used to conduct our research on. First of all we show some general statistics of the (L)ETFs and their options. Next we show the necessity of a sophisticated comparison method by showing the plots of the IV both with and without moneyness scaling. Lastly we discuss our filtering process and present the summary statistics of the filtered dataset.

3.1 Data Selection

For this research we analyze a set of (L)ETFs that track the S&P 500. This set consists of the regular SPDR S&P500 (SPY) ETF and the following four LETFs: ProShares Ultra S&P500 (SSO), ProShares UltraShort S&P500 (SDS), ProShares UltraPro S&P500 (UPRO) and ProShares UltraPro Short S&P500 (SPXU), with betas of +2, -2, +3 and -3 respectively. From now on the given abbreviations will be used when referring to these (L)ETFs. The data of the American options is retrieved from OptionMetrics and contains the following variables: closing bid & ask quotes, strike prices, underlying asset prices, expiration dates, implied volatilities, daily trading volume of the options and daily trading volume of the underlying (L)ETF. The contracts that have missing values for the implied volatility have been excluded. The UPRO and SPXU LETFs became available on June 26th 2009, however the options on these LETFs were not released until July 6th 2009. Therefore the dataset stretches from July 6th 2009 to April 29th 2016, consisting of 1718 trading days. The SPY, SSO and UPRO (L)ETFs have option data on all 1718 trading days. For SDS options the data for Oct 5th 2012 is missing, while SPXU options misses data on May 11th 2012 and Jan 24th 2014. To ensure comparability between the datasets we removed these data points from all datasets and therefore we end up with 1715 trading days. The expense fees of the LETFs were not available on OptionMetrics, and therefore we retrieved them from ETF Report (ETF.com). As a proxy for the risk-free rate we use the US treasury yields for which we retrieved the daily data for 1, 3 and 6 months and 1, 2 years from Quandl. Lastly we note that it is common practice for LETFs to (reverse) split its number of shares when the price has reached a certain level, which is done to ensure that its price will remain in a convenient trading range. For instance a split of 2:1 means that on that specific date every share holder receives double the number of shares, causing a 50% decrease of its price. Once such a split is announced, the corresponding options are altered in such a way that it has neither a positive or negative effect for its holders. In appendix D the split history of our LETF shares can be found.

3.2 General Statistics of (L)ETFs and their Options

The chosen (L)ETFs with their corresponding theoretical betas, annual fees and their average daily trading volumes are presented in table 1. The annual fees are paid by the fund holder to the issuer. In the same table we present some general information of the available options for our (L)ETFs. First of all the average number of available option contracts on each trading day is given. These option contracts differ on both moneyness and time-to-maturity. We also show the average of the aggregated trading volume of all available options on each trading day. We first aggregate the total trading volume of all available option contracts per day and then took the average of this. The volume is defined as the number of (L)ETF option contracts that are traded per day (both buying and selling). It can be seen that options on the SPY ETF have a significantly higher liquidity than the options on the LETFs. Furthermore, the SPY ETF has more different option contracts available than any of the LETFs. Therefore the data of the options on the SPY ETF contains a lot more and richer information about the IVS than its leveraged counterparts. It also stands out that the SDS and SPXU ETFs and their options have a higher liquidity than their positive counterparts, SSO and UPRO correspondingly.

In appendix E we present how both the daily aggregated trading volume of options on (L)ETFs and the number of available contracts have changed over the timespan of our dataset. The most important observation in regards to the trading volume is that the trading volume on SPY, UPRO and SPXU options grew rapidly until 2011, whereafter it stayed relatively on the same level. For both SSO and SDS options the trading volume grew until 2011 and declined thereafter. For the number of available contracts we conclude that the number of available option contracts for the SPY ETF has grown steadily over time, from 2101 on July 6th 2009 to 3731 on April 29th

2016. Furthermore, the number of available option contracts for SSO and UPRO has also grown, whereas the number of available option contracts for SDS and SPXU maintained relatively on the same level. Interestingly enough, despite the growth of available option contracts we find that the aggregated trading volume on each trading day did not grow over this period for any of the (L)ETF options.

Table 1: The set of (L)ETFs and their options that track the S&P 500.

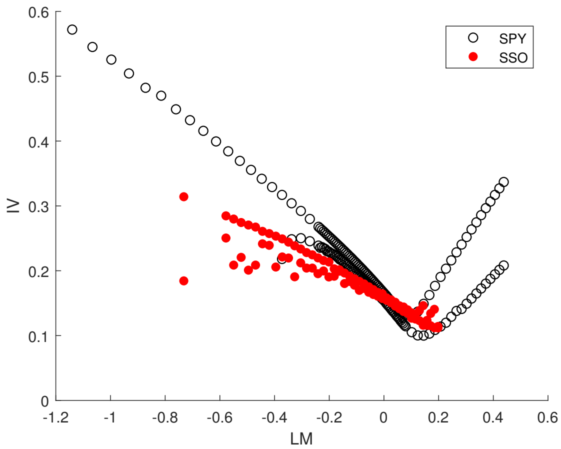
	SPY	**SSO**	**SDS**	**UPRO**	**SPXU**
Theoretical Beta (β)	+1	+2	−2	+3	−3
Annual Fee (c)	0.09%	0.90%	0.89%	0.95%	0.90%
Average Trading Volume ETFs	155,100,000	10,761,000	20,416,000	2,162,400	6,003,400
Average Number of Available Option Contracts	2541	680	483	611	330
Average Daily Aggregated Trading Volume Options	2,072,000	12,848	32,487	1,161	2,274

Notes: This table contains the different (L)ETFs with their corresponding theoretical betas, annual fees and the average daily trading volumes. Next the average number of available option contracts (differing over time-to-maturity and strike price) are presented. Lastly the average of the aggregated trading volumes of all available options on each trading day is given. The data is based on 1715 trading days spanning July 6th 2009 to April 29th 2016.

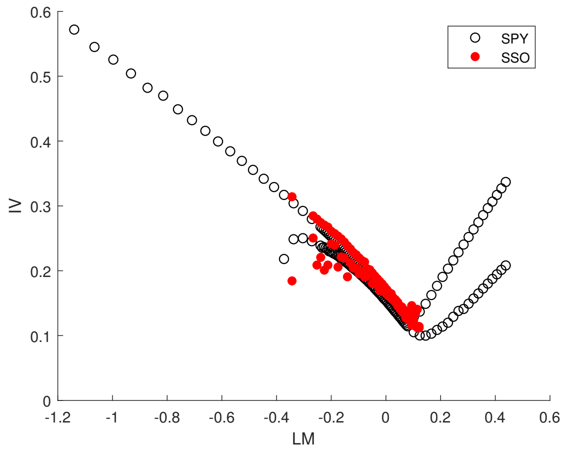
3.3 Comparison of (L)ETF options

Our goal is to leverage the data of ETF options for the purpose of modeling the IVS of LETF options. Before we are able to do so we show the necessity of a sophisticated comparison method. For this purpose we present the plots of the implied volatility smiles of the LETF options in comparison to the SPY options. The option data is arbitrarily chosen; we chose March 28 2016 for which we selected all options with 172 days to maturity. It has to be noted that we brought the IV of the LETF options on the same scale as the SPY options, by means of equation 1 from the Methods section. When applying the approximate moneyness scaling method by Leung & Sircar (2015) from equation 3 on the same selected options we retrieve the moneyness scaled variants. Both plots can be seen in figure 2. It can be seen that for all LETFs the IV smile structures are now on the same level as the IV smile structure of the SPY options. However, it can also be seen that there still are some discrepancies left in the IV of LETF options. Furthermore, it is clear that the data on ETF options contains more data points and seems to be more consistent than the data on LETF options. Lastly, it can be seen that LETF options are not spread out over the same moneyness interval as the ETF options.

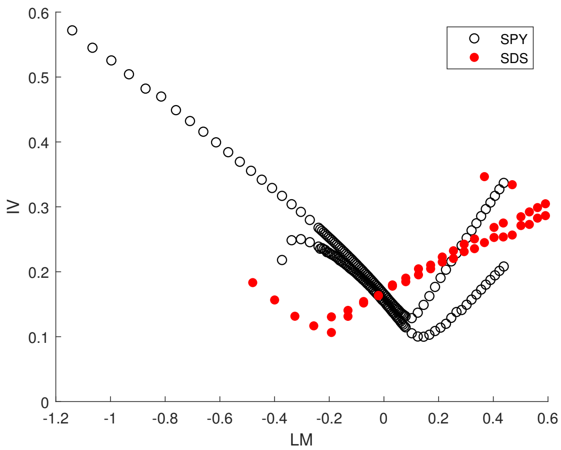
SSO (+2) - Before Moneyness Scaling



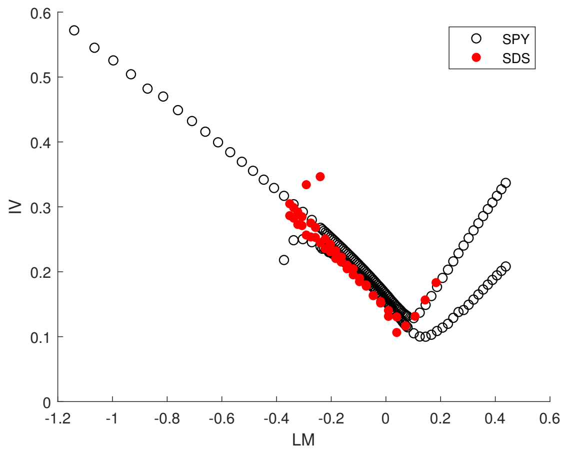
SSO (+2) - After Moneyness Scaling



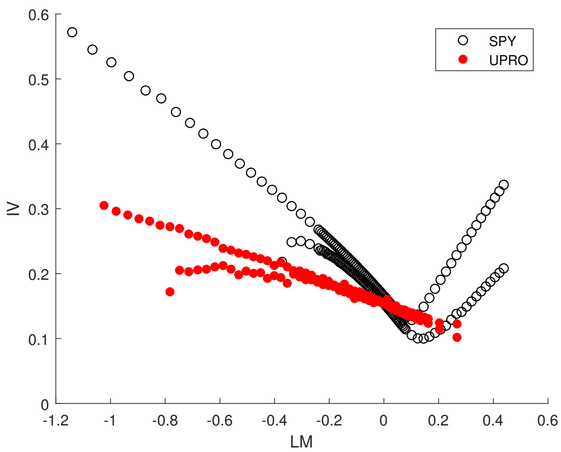
SDS (-2) - Before Moneyness Scaling



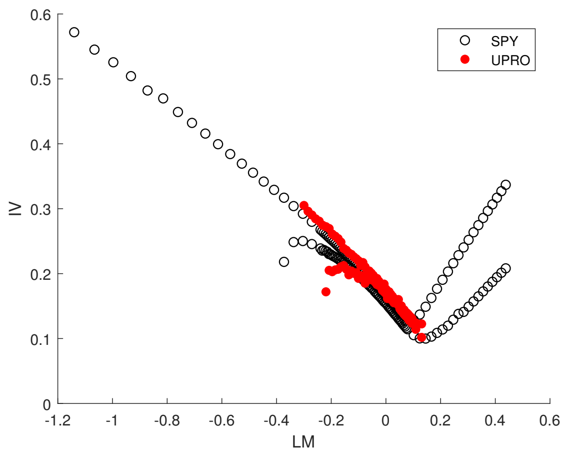
SDS (-2) - After Moneyness Scaling



UPRO (+3) - Before Moneyness Scaling



UPRO (+3) - After Moneyness Scaling



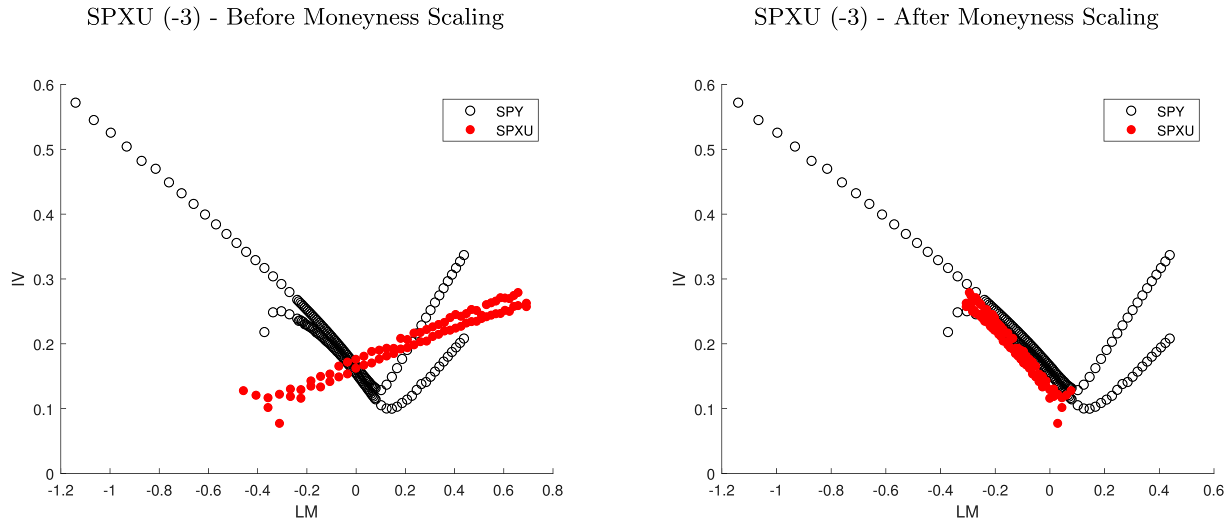


Figure 2: The SPY and LETF implied volatility smiles before and after moneyness scaling. After moneyness scaling both the IV and the moneyness are on the scale of the SPY options. All options have 172 days to maturity, whereas the data is from 28 March 2016. All IVs are on the same scale as the SPY IV.

3.4 Filtering of the Dataset

To enhance the data quality for the purpose of modeling the IVS, we filter out observations that are not likely to represent traded prices in well-functioning and liquid option markets. Unfortunately the empirical sections of both Leung & Sircar (2015) and Härdle et al. (2016) are limited, and therefore they do not provide any filters for the analysis of LETF option data. Consequently we are obliged to perform our own empirical analysis to combine filters from different academic papers that study the IVS of options.

As shown in the previous subsection we can use the moneyness scaling approach to get the LETF data on the same scale as the ETF data, and thus we are now able to construct a general framework of filters. An empirical analysis has been performed by checking which existing filters find a good balance between eliminating noisy data and leaving enough data to conduct our research on. In appendix H the histograms of each (L)ETF for both scaled IV and scaled moneyness are presented. Based on the available literature and our own empirical research we construct five filters. First of all we follow the same reasoning as Bernales & Guidolin (2014) and van der Wel et al. (2016) and exclude contracts with less than 10 days to expiry or with more than 1 year to expiry. Secondly we exclude contracts that have a price lower than \$0.05, which is a filter that is also used by van der Wel et al. (2016). Thirdly van der Wel et al. (2016) exclude options that have an IV of over 0.7. Due to the risky characteristics of the options on LETFs many contracts have an IV that is much higher. Therefore we first bring all the implied volatilities on the same level as the SPY options and then apply the filter on these scaled IVs. Fourthly we follow Cont et al. (2002) and exclude contracts that have extreme moneyness values, thus all contracts that have a moneyness smaller than 0.5 or larger than 1.5. To ensure that this filter can be used on the LETFs we apply it to the scaled moneyness. Lastly, due to the lack of liquidity on the option market of LETFs a substantial amount of options have a best bid quote of 0 on a certain day, whereas especially the option markets of UPRO and SPXU LETFs are particularly illiquid. In those cases it seems unfair to take the midpoint between the bid and the ask quotes as the price of the option. Leung & Santoli (2016) deal with this by excluding options that have a bid quote of less than 0.5, however this seems to eliminate too many observations from our dataset. Therefore we only drop the contracts that have a bid quote equal to 0 on a certain day. To show how each of our chosen filters impacts the final number of observations we present a table with the number of lost observations per filter in appendix G.

In table 2 the summary statistics of the (scaled) IVs per (L)ETF based on the filtered dataset is presented. The (scaled) moneyness dimension is divided in four groups, whereas the maturity dimension is divided in three groups. The trading frequency shows the percentage of trading days in which we observe at least one trade for an option contract with the corresponding moneyness and time-to-maturity characteristics. Furthermore the table presents the average (scaled) IV per group with its corresponding standard deviation in the brackets next to it. It can be seen that the standard deviation of the IV is relatively high for out-of-the-money and deep in-the-money options. Furthermore the standard deviation of the IV is higher for short-term than for long-term options. In terms of the trading frequency it can be seen that the SPY options are very liquid, particularly when compared to the deep out-of-the-money and deep in-the-money LETF options. The distribution of our filtered dataset over the different segments of the IVS can be seen in appendix H. The main take-aways are that the deep in-the-money options cover a relatively small part of the dataset. Furthermore the distribution gets less dense as the time to maturity increases.

Table 2: Summary Statistics of the (scaled) IVs of options on (L)ETFs.

	Short Term (10-90 days)		**Medium Term (90-180 days)**		**Long Term (180-365 days)**	
	Trading Freq. (%)	**Mean $\widetilde{IV}$ (%)**	**Trading Freq. (%)**	**Mean $\widetilde{IV}$ (%)**	**Trading Freq. (%)**	**Mean $\widetilde{IV}$ (%)**
SPY (+1)						
$0.5 \leq M \leq 0.9$	100.0	34.61 (11.4)	100.0	30.49 (8.6)	100.0	27.53 (6.4)
$0.9 \leq M \leq 1$	100.0	18.65 (5.3)	100.0	19.29 (4.5)	100.0	19.76 (4.1)
$1.0 \leq M \leq 1.1$	100.0	13.73 (4.5)	100.0	15.10 (4.4)	100.0	17.03 (4.1)
$1.1 \leq M \leq 1.5$	84.4	26.27 (11.3)	97.8	17.32 (5.3)	99.9	16.45 (4.1)
SSO (+2)						
$0.5 \leq \widetilde{M} \leq 0.9$	91.3	33.28 (10.9)	84.0	27.31 (7.4)	54.0	25.13 (6.0)
$0.9 \leq \widetilde{M} \leq 1$	100.0	18.22 (5.7)	95.9	19.36 (4.7)	62.3	19.75 (4.1)
$1.0 \leq \widetilde{M} \leq 1.1$	100.0	13.97 (4.9)	98.7	15.68 (4.5)	73.4	16.55 (4.0)
$1.1 \leq \widetilde{M} \leq 1.5$	31.6	25.45 (10.6)	61.3	16.84 (5.2)	47.9	16.09 (4.4)
SDS (-2)						
$0.5 \leq \widetilde{M} \leq 0.9$	98.0	32.94 (10.5)	99.0	28.36 (6.9)	81.6	25.19 (5.7)
$0.9 \leq \widetilde{M} \leq 1$	100.0	19.57 (6.5)	99.4	19.75 (5.5)	80.8	18.63 (4.9)
$1.0 \leq \widetilde{M} \leq 1.1$	100.0	16.20 (6.1)	91.6	16.74 (5.4)	52.7	16.37 (5.1)
$1.1 \leq \widetilde{M} \leq 1.5$	28.5	36.46 (12.9)	23.7	22.95 (6.8)	15.0	20.65 (5.7)
UPRO (+3)						
$0.5 \leq \widetilde{M} \leq 0.9$	61.6	30.95 (9.9)	50.6	24.65 (5.35)	30.4	22.89 (4.5)
$0.9 \leq \widetilde{M} \leq 1$	99.7	18.60 (5.4)	69.0	19.08 (3.95)	36.3	19.99 (4.2)
$1.0 \leq \widetilde{M} \leq 1.1$	99.9	14.88 (4.9)	87.4	16.11 (3.93)	53.8	17.10 (4.0)
$1.1 \leq \widetilde{M} \leq 1.5$	18.5	23.45 (9.5)	48.4	15.89 (4.25)	43.4	15.43 (3.9)
SPXU (-3)						
$0.5 \leq \widetilde{M} \leq 0.9$	78.3	30.02 (9.85)	85.7	24.91 (6.1)	61.3	22.38 (4.7)
$0.9 \leq \widetilde{M} \leq 1$	100.0	19.86 (6.4)	91.1	18.19 (5.0)	47.4	17.41 (4.4)
$1.0 \leq \widetilde{M} \leq 1.1$	96.7	17.01 (6.3)	39.2	16.15 (5.0)	10.6	16.01 (4.4)
$1.1 \leq \widetilde{M} \leq 1.5$	4.8	35.96 (12.9)	2.7	24.84 (8.6)	1.1	22.76 (7.3)

Notes: This table contains summary statistics for the implied volatilities across moneyness ($\widetilde{M}$) and time-to-maturity (in calendar days to expiration). For the LETFs both the IV and the moneyness coordinate are transformed to the same scale as the SPY options. The trading frequency shows the percentage of trading days in which we observe at least one trade for an option contract with the corresponding moneyness and time-to-maturity characteristics. Furthermore the mean $\widetilde{IV}$ shows the average of the IV for that particular subgroup and has its standard deviation shown in the brackets next to it. The data covers the period between July 6th 2009 to April 29th 2016.

4. EMPIRICAL ANALYSIS

In this section we present the results of our empirical analysis. We first describe how we modeled the IVS of LETF options and present the in-sample results. Next we assess the out-of-sample forecast performance both in a statistical and economical setting.

4.1 Modeling the IVS of LETF Options

We use the Leung & Sircar method, as described in the Methods section, to model the IVS for each day in our dataset. First we estimate equation 4 and use these estimates to calibrate the stochastic volatility group parameters using equation 8. For the yearly risk-free rate we use the 1 year US treasury yield rate. Finally, the stochastic volatility group parameters are used to find the estimates that are needed in equation 6 leaving us with a model for the IVS of the LETF options per day. In figure 3 we plotted the modeled IVS along with the real IV data points on the arbitrarily chosen date of 12 May 2015. On this date we have 932 IV values for the SSO options, 328 for the SDS options, 1078 for the UPRO options and 253 for the SPXU options. The Leung & Sircar method seems to be able to capture the data points quite well. However, as can clearly be seen in the plot of the SDS options, the used deterministic function is only able to reproduce the implied volatility skew (asymmetric smile) and not the implied volatility smile.

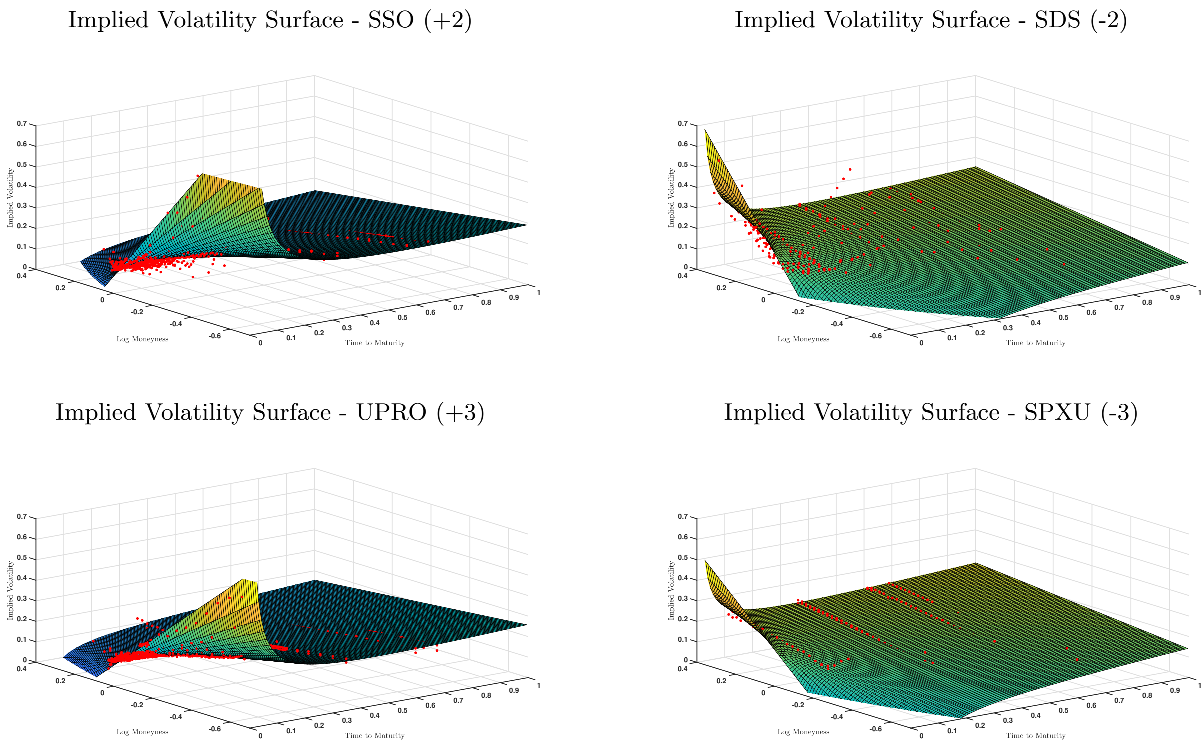


Figure 3: The IVS as modeled using the Leung & Sircar Method for all LETF options on 12 May 2015.

Before we analyze the forecasting performance of the Leung & Sircar method we evaluate its in-sample fit on the real IV data. The in-sample performance is assessed by using the in-sample variants of equations 17 and 18 from the Methods section, the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) respectively. The MCPDC is not considered, as this measure is used to assess the dynamic aspect of the modeling. Furthermore, it is interesting to see how this fit compares to the fit of the model that is directly estimated on the option data of the LETF options. Therefore we also present the results of the direct estimation method (DE, benchmark 1). The other benchmark models that are presented in the Methods section are based on a dynamic approach and therefore will not be assessed for the in-sample fit.

The in-sample results are shown in table 3. As can be seen the Leung & Sircar method performs significantly worse than the direct estimation approach, for all LETF options. Yet this is of no surprise, as the LETF option data is also the real data that we are comparing against. For both methods we use the same deterministic function to model the IVS, however the difference lies in the estimation of its coefficients. Since we are using the ETF option data as a proxy to capture the information of the LETF IVS it would be hard to beat a method that directly estimates the coefficients on the same dataset as the one that we are comparing against. We are hopeful though, that the Leung & Sircar method proves more useful when forecasting the IVS. The idea is based on the assumption that the IV values of the LETF options should convert towards their 'true' values, which should be on the same level as the information-rich and highly liquid SPY options.

Table 3: Overall In-Sample Performance

	SSO (+2)	**SDS (-2)**	**UPRO (+3)**	**SPXU (-3)**
Panel 1: RMSE				
LS	0.053	0.071	0.046	0.054
DE	0.040	0.050	0.027	0.035
Panel 2: MAE				
LS	0.037	0.042	0.034	0.035
DE	0.029	0.034	0.020	0.025

Notes: This table contains the in-sample fit, presented in two different measures. The Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE). The used models are the Leung & Sircar method (LS) and the Direct Estimation on LETF options (DE, benchmark 1).

Before we move on towards the out-of-sample performance we investigate how the Leung & Sircar method performs over different segments of the IVS. Therefore we calculate the same measures over different segments of the IVS. The scaled moneyness dimension is divided in four groups, whereas the maturity dimension is divided in three groups. The results for the RMSE are shown in tables 4, whereas the results for the MAE are shown in appendix I. We find that for the SSO and SDS options there are several segments in which the LS method performs better than the DE benchmark. For the SSO options this is the case for the long-term options, whereas for the SDS options the Leung & Sircar method outperforms the direct estimation method in all cases that are close to at-the-money (a moneyness between 0.9 and 1.1).

4.2 Out-of-Sample Forecast Performance

Next we assess the dynamic performance of the Leung & Sircar method by out-of-sample forecasting the IVS. We use a recursive moving-window approach to find the one-day-ahead predictions and use these forecasts to systematically evaluate and compare the out-of-sample performance of all models. This is done using a six-month rolling window of data (252/2 = 126 days). Over this sample we discard the first and the last forecast. The first forecast, because for the MCPDC measure we do not have a previous forecast for that point in time. And the last forecast due to the lack of real data to calculate the performance measures. This leaves us with a total of 1,589 days for which we have forecasted values of the IVS. (From 4 Jan 2010 until 29 Apr 2016)

Table 4: In-Sample Performance per Segment - RMSE.

	Short Term (10-90 days)		**Medium Term (90-180 days)**		**Long Term (180-365 days)**	
	LS	**DE**	**LS**	**DE**	**LS**	**DE**
SSO (+2)						
$0.5 \leq \widetilde{M} \leq 0.9$	0.062	0.045	0.029	0.027	0.027	0.028
$0.9 \leq \widetilde{M} \leq 1$	0.043	0.038	0.022	0.017	0.016	0.019
$1.0 \leq \widetilde{M} \leq 1.1$	0.054	0.043	0.027	0.019	0.017	0.018
$1.1 \leq \widetilde{M} \leq 1.5$	0.083	0.068	0.036	0.029	0.023	0.026
SDS (-2)						
$0.5 \leq \widetilde{M} \leq 0.9$	0.158	0.109	0.059	0.048	0.040	0.036
$0.9 \leq \widetilde{M} \leq 1$	0.052	0.056	0.026	0.030	0.017	0.021
$1.0 \leq \widetilde{M} \leq 1.1$	0.039	0.042	0.019	0.022	0.015	0.017
$1.1 \leq \widetilde{M} \leq 1.5$	0.054	0.046	0.019	0.017	0.019	0.016
UPRO (+3)						
$0.5 \leq \widetilde{M} \leq 0.9$	0.049	0.033	0.023	0.019	0.021	0.017
$0.9 \leq \widetilde{M} \leq 1$	0.044	0.021	0.024	0.012	0.017	0.012
$1.0 \leq \widetilde{M} \leq 1.1$	0.053	0.024	0.027	0.012	0.017	0.012
$1.1 \leq \widetilde{M} \leq 1.5$	0.064	0.038	0.032	0.018	0.019	0.014
SPXU (-3)						
$0.5 \leq \widetilde{M} \leq 0.9$	0.097	0.069	0.042	0.031	0.030	0.025
$0.9 \leq \widetilde{M} \leq 1$	0.051	0.040	0.030	0.021	0.021	0.018
$1.0 \leq \widetilde{M} \leq 1.1$	0.041	0.032	0.025	0.017	0.019	0.016
$1.1 \leq \widetilde{M} \leq 1.5$	0.042	0.033	0.021	0.016	0.019	0.016

Notes: This table contains the in-sample fit per segment as measured in the Root Mean Squared Error (RMSE). The used models are the Leung & Sircar method (LS) and the Direct Estimation on LETF options (DE, benchmark 1). The moneyness of the LETF options is transformed to the same scale as the SPY options.

In table 5 we present the results of the three evaluation methods for the different LETF options. We calculated these measures for each day and then took the average over these results. This has to be done per day instead of averaging over the whole dataset, because the number of contracts increases over time (at least for the SSO and UPRO options) and therefore this would assign more weight on the observations on the later days. Furthermore it has to be noted that we did not calculate the MCPDC for the random walk model, as it requires that an individual option is in the sample for the three consecutive days. The used models are the Leung & Sircar method (LS), direct estimation on LETF options (DE, benchmark 1), the Bernales & Guidolin method (BG, benchmark 2), the random walk (RW, benchmark 3) and the Strawman (SM, benchmark 4). We have to note that the random walk method does not produce forecasts for every single IV point, due to the fact that for some options there is no data available of the previous day. Therefore, the random walk method has less evaluation points than the others. As can be seen none of the models are able to beat the random walk forecasts. A possible explanation for this could be the fact that the used deterministic function to model the IVS is unable to fully capture the volatility smile structure of the IVS. The emboldened values indicate the best performing models when disregarding the random walk forecasts, which is common practice in literature on the IVS. As can be seen the Leung & Sircar method is only able to beat the benchmark models once, on the basis of the MAPE for the SDS options. In all other cases the direct estimation method outperforms the other models, closely followed by the BG method. It seems that the extra information on the SPY options that is included as exogenous variables in the Bernales & Guidolin method causes more estimation uncertainty than it is able to provide in extra forecasting power.

Table 5: Overall Out-of-Sample Forecast Performance

	LS	**DE**	**BG**	**RW**	**SM**
SSO (+2)					
RMSPE	0.043	**0.041**	0.041	0.025	0.048
MAPE	0.032	**0.030**	0.030	0.015	0.035
MCPDC (%)	54.82	**55.46**	55.33	NA	54.30
SDS (-2)					
RMSPE	0.053	**0.051**	0.052	0.031	0.057
MAPE	**0.033**	0.034	0.034	0.017	0.036
MCPDC (%)	55.82	**55.83**	55.61	NA	55.17
UPRO (+3)					
RMSPE	0.040	**0.029**	0.030	0.021	0.043
MAPE	0.031	**0.021**	0.022	0.014	0.033
MCPDC (%)	54.93	**57.07**	56.97	NA	54.53
SPXU (-3)					
RMSPE	0.042	**0.038**	0.039	0.028	0.046
MAPE	0.029	**0.026**	0.026	0.018	0.031
MCPDC (%)	58.87	**60.01**	59.97	NA	58.19

Notes: This table contains the out-of-sample forecast performance, presented in three different measures. The Root Mean Squared Prediction Error (RMSPE), the Mean Absolute Prediction Error (MAPE) and the Mean Correct Prediction of Direction of Change (MCPDC). The used models are the Leung & Sircar method (LS), direct estimation on LETF options (DE, benchmark 1), the Bernales & Guidolin method (BG, benchmark 2), the random walk (RW, benchmark 3) and the Strawman (SM, benchmark 4). The moneyness of the LETF options is transformed to the same scale as the SPY options. For each LETF option set we emboldened the best performing model. It is important to note that we neglected the results of the random walk. Furthermore the MCPDC is not calculated for the random walk model, as this requires that an individual option is in the sample for the three consecutive days. We used a six month rolling window of data to find the one-day-ahead predictions.

Since the Leung & Sircar method is not able to outperform our benchmark models (except for the aforementioned case) it is interesting to see whether the method could still be useful for particular segments of the IVS. Therefore we calculated the same prediction measures over different segments of the IVS, just like we did for the in-sample evaluation. The results for the RMSPE and for the MCPDC can be seen in tables 6 and 7 respectively. The results for the MAPE can be found in appendix J. For each LETF option set we emboldened the best performing model in its particular segment. Once again we have to note that we neglect the results of the random walk (RW). Furthermore it has to be noted that the MCPDC is not calculated for the random walk model, as this requires that an individual option is in the sample for the three consecutive days. These results seem more promising, as the Leung & Sircar method is actually able to outperform the benchmark models in several segments. Particularly for the SDS options the Leung & Sircar method has a much better performance for all three performance measures. Furthermore the performance on the long-term SSO options is better than the performance of the other benchmark models. Unfortunately its performance on the UPRO and SPXU options is still rather poor. In most of these cases the direct estimation method has the best performance, whereas in some cases the Bernales and Guidolin method has a better performance.

Table 6: Out-of-Sample Forecast Performance per Segment - RMSPE.

	Short Term (10-90 days)					Medium Term (90-180 days)					Long Term (180-365 days)				
	LS	**DE**	**BG**	**RW**	**SM**	**LS**	**DE**	**BG**	**RW**	**SM**	**LS**	**RW**	**BG**	**DE**	**SM**
SSO (+2)															
$0.5 \leq \widetilde{M} \leq 0.9$	0.046	**0.043**	0.043	0.032	0.053	0.028	**0.028**	0.028	0.024	0.029	**0.024**	0.029	0.030	0.018	0.027
$0.9 \leq \widetilde{M} \leq 1.0$	0.042	0.041	**0.041**	0.015	0.043	0.022	0.018	**0.017**	0.008	0.023	**0.016**	0.021	0.022	0.009	0.017
$1.0 \leq \widetilde{M} \leq 1.1$	0.049	0.044	**0.044**	0.015	0.054	0.026	0.019	**0.018**	0.008	0.027	**0.017**	0.020	0.021	0.008	0.018
$1.1 \leq \widetilde{M} \leq 1.5$	**0.067**	0.068	0.068	0.032	0.077	0.032	0.028	**0.028**	0.017	0.035	**0.022**	0.028	0.028	0.012	0.023
SDS (-2)															
$0.5 \leq \widetilde{M} \leq 0.9$	0.124	0.110	**0.108**	0.048	0.128	**0.048**	0.048	0.048	0.025	0.050	0.034	**0.032**	0.033	0.024	0.035
$0.9 \leq \widetilde{M} \leq 1.0$	**0.047**	0.057	0.057	0.015	0.052	**0.025**	0.030	0.030	0.007	0.027	**0.016**	0.017	0.018	0.008	0.017
$1.0 \leq \widetilde{M} \leq 1.1$	**0.039**	0.045	0.044	0.014	0.039	**0.019**	0.022	0.022	0.006	0.020	**0.014**	0.015	0.015	0.007	0.015
$1.1 \leq \widetilde{M} \leq 1.5$	**0.043**	0.045	0.046	0.034	0.050	0.020	0.020	**0.019**	0.016	0.021	0.017	**0.016**	0.016	0.013	0.020
UPRO (+3)															
$0.5 \leq \widetilde{M} \leq 0.9$	0.043	**0.035**	0.036	0.031	0.046	0.023	**0.021**	0.022	0.016	0.023	0.020	**0.020**	0.020	0.011	0.022
$0.9 \leq \widetilde{M} \leq 1.0$	0.044	0.025	**0.024**	0.016	0.044	0.023	**0.013**	0.013	0.009	0.023	0.016	**0.014**	0.014	0.007	0.017
$1.0 \leq \widetilde{M} \leq 1.1$	0.049	**0.025**	0.025	0.016	0.053	0.026	0.013	**0.013**	0.008	0.027	0.017	0.014	**0.014**	0.007	0.017
$1.1 \leq \widetilde{M} \leq 1.5$	0.053	0.042	**0.042**	0.026	0.062	0.029	**0.018**	0.019	0.011	0.032	0.019	0.016	**0.016**	0.007	0.020
SPXU (-3)															
$0.5 \leq \widetilde{M} \leq 0.9$	0.076	0.038	0.072	**0.071**	0.081	0.038	0.033	**0.032**	0.023	0.040	0.027	**0.024**	0.026	0.022	0.028
$0.9 \leq \widetilde{M} \leq 1.0$	0.047	0.041	**0.041**	0.017	0.050	0.028	0.022	**0.022**	0.013	0.029	0.020	**0.017**	0.018	0.014	0.020
$1.0 \leq \widetilde{M} \leq 1.1$	0.040	0.035	**0.035**	0.015	0.041	0.024	0.025	0.018	**0.018**	0.012	0.018	**0.016**	0.016	0.014	0.019
$1.1 \leq \widetilde{M} \leq 1.5$	0.039	**0.036**	0.036	0.033	0.042	0.021	0.018	**0.017**	0.017	0.022	0.017	**0.016**	0.017	0.014	0.019

Notes: This table contains the out-of-sample forecast performance per segment as measured in the Root Mean Squared Prediction Error (RMSPE). The used models are the Leung & Sircar method (LS), direct estimation on LETF options (DE, benchmark 1), the Bernales & Guidolin method (BG, benchmark 2), the random walk (RW, benchmark 3) and the Strawman (SM, benchmark 4). The moneyness of the LETF options is transformed to the same scale as the SPY options. For each segment we emboldened the best performing model. It is important to note that we neglected the results of the random walk. We used a six month rolling window of data to find the one-day-ahead predictions.

Table 7: Out-of-Sample Forecast Performance per Segment - MCPDC (%).

	Short Term (10-90 days)					Medium Term (90-180 days)					Long Term (180-365 days)				
	LS	**DE**	**BG**	**RW**	**SM**	**LS**	**DE**	**BG**	**RW**	**SM**	**LS**	**RW**	**BG**	**DE**	**SM**
SSO (+2)															
$0.5 \leq \widetilde{M} \leq 0.9$	54.83	**57.10**	56.68	NA	54.29	56.55	57.21	**57.45**	NA	56.89	58.00	**55.91**	56.40	NA	57.41
$0.9 \leq \widetilde{M} \leq 1.0$	51.29	51.70	**51.88**	NA	51.10	53.11	54.29	**54.49**	NA	51.88	**56.82**	53.48	53.65	NA	56.75
$1.0 \leq \widetilde{M} \leq 1.1$	48.99	**50.68**	50.63	NA	48.29	52.18	54.78	**55.35**	NA	51.27	56.02	54.15	**54.21**	NA	55.71
$1.1 \leq \widetilde{M} \leq 1.5$	**56.15**	55.14	54.91	NA	53.20	56.53	57.83	**57.94**	NA	54.85	**58.00**	55.10	55.30	NA	57.41
SDS (-2)															
$0.5 \leq \widetilde{M} \leq 0.9$	54.85	55.14	**55.39**	NA	53.21	**54.84**	53.83	53.73	NA	54.27	**57.99**	56.99	56.85	NA	57.90
$0.9 \leq \widetilde{M} \leq 1.0$	**50.15**	49.09	49.01	NA	49.96	**51.14**	50.09	50.41	NA	51.04	**55.33**	53.20	53.54	NA	54.67
$1.0 \leq \widetilde{M} \leq 1.1$	52.32	51.25	51.07	NA	**52.62**	**52.48**	50.82	50.93	NA	51.85	**55.74**	53.12	52.98	NA	54.39
$1.1 \leq \widetilde{M} \leq 1.5$	58.25	**58.70**	58.07	NA	56.75	57.16	56.94	**57.58**	NA	57.29	56.67	**57.20**	56.72	NA	54.62
UPRO (+3)															
$0.5 \leq \widetilde{M} \leq 0.9$	57.61	**59.30**	58.62	NA	56.89	57.11	56.81	56.95	NA	**57.25**	55.73	54.45	54.55	NA	**56.07**
$0.9 \leq \widetilde{M} \leq 1.0$	52.02	**54.70**	55.00	NA	52.39	54.83	56.91	**58.13**	NA	53.78	**55.95**	54.88	54.51	NA	55.92
$1.0 \leq \widetilde{M} \leq 1.1$	50.34	**55.04**	54.76	NA	50.81	53.18	58.54	**59.16**	NA	52.67	55.18	**55.84**	55.21	NA	55.02
$1.1 \leq \widetilde{M} \leq 1.5$	55.34	**58.46**	57.82	NA	52.47	53.63	57.83	**58.24**	NA	51.84	55.41	**56.07**	55.90	NA	55.10
SPXU (-3)															
$0.5 \leq \widetilde{M} \leq 0.9$	55.81	**56.32**	56.04	NA	54.90	54.40	56.62	**56.84**	NA	54.09	58.68	**60.00**	59.38	NA	58.18
$0.9 \leq \widetilde{M} \leq 1.0$	50.82	51.01	50.98	NA	**51.13**	54.41	**56.14**	55.79	NA	53.80	58.83	**60.51**	60.27	NA	57.57
$1.0 \leq \widetilde{M} \leq 1.1$	**52.30**	52.05	52.12	NA	52.89	56.07	57.81	**58.46**	NA	55.78	59.26	61.29	**61.66**	NA	58.84
$1.1 \leq \widetilde{M} \leq 1.5$	60.09	60.63	**60.70**	NA	59.49	60.18	63.15	**63.73**	NA	59.91	60.67	**61.33**	61.18	NA	59.40

Notes: This table contains the out-of-sample forecast performance per segment as measured in the Mean Correct Prediction of Direction of Change (MCPDC). The used models are the Leung & Sircar method (LS), direct estimation on LETF options (DE, benchmark 1), the Bernales & Guidolin method (BG, benchmark 2), the random walk (RW, benchmark 3) and the Strawman (SM, benchmark 4). The moneyness of the LETF options is transformed to the same scale as the SPY options. For each segment we emboldened the best performing model. It is important to note that for the MCPDC measure we did not include the random walk model, as this requires that an individual option is in the sample for the three consecutive days. We used a six month rolling window of data to find the one-day-ahead predictions.

4.3 Economical Evaluation

To analyze if the forecast performance of our models can be utilized in practice we evaluate them in economical sense by backtesting the trading strategy as presented in the Methods section. This trading strategy is characterized by its simplicity and therefore has to be interpreted as a lower bound for the actual profits, as an experienced trading desk will be able to optimize this strategy in various ways. We note that for our evaluation we do not take trading fees into account. Moreover we simply use the mid-prices for our option prices, not taking into account the size of the quoted spreads and their corresponding effective spreads. To keep the trading strategy simple we solely allow it to use call options and thus disregard all put options. Furthermore we remove all option contracts from the dataset that have no data available for the next day. We also lookup the trading days on which a (reverse) split occurred for the shares of the LETFs and exclude the trading day before. This is done as this could cause difficulties when calculating the difference between both the option prices and the price of the underlying on the day of the stock split and the day before.

As discussed in the Methods section our trading strategy is analyzing IV strings separately and provides a signal for this given IV string. To keep the trading strategy as simple as possible we only evaluate one IV string per day. Therefore we have to decide beforehand whether this IV string comes from the short-term, medium-term or long-term option contracts. In the Data and Preliminary Results section we presented summary statistics of our option data set and concluded that the IV of short-term options have a relatively high standard deviation, whereas long-term options have a relatively low standard deviation of its IV. We compromise and use the medium term option contracts, and thus use 90 days-to-expiration as our cut-off-point. This ensures that our algorithm selects the first IV string it finds that has more than 90 days-to-expiration. As explained in the Methods section there are three cases that can occur for each string. The riskless interest rate over one day that is needed when case 3 occurs is retrieved from the US treasury yields. In figure 4 we present each of these cases in graphical form for an easier understanding. Both the strings and dates are arbitrarily chosen, purely for presentational purpose. The black dots are the real data points on time t, whereas the blue line represents the forecasted values for time $t+1$ (as modeled by the Leung & Sircar method). The option that is selected by the strategy is marked in a different color and is simply the one that deviates the most from its forecasted value, as we expect the real values to convert towards their forecasted values.

As proposed in the Methods section we planned to invest an initial amount of $C_0 = \$1000$ per day in our selected portfolios. However, when backtesting this strategy without any restrictions we found that when the construction cost of a portfolio is close to zero it allows the algorithm to long or short a very large quantity of portfolios on one day (e.g. at a certain date the strategy, based on the modeling of the IV of SSO options using the Leung & Sircar method, favored a portfolio with a net construction cost of \$0.004, allowing the algorithm to buy 250,480 of these portfolios). This is undesirable as this would require a very large margin account, in comparison to the initial investment amount, to even be allowed to short such an amount of options. It has to be noted that it is rather strange that Bernales & Guidolin (2014) do not address this problem, as it can be seen that their standard deviations of the profit are all relatively high, and thus that they are dealing with the same kind of problem. We decided to restrict the algorithm from taking any action when the the net construction cost of a portfolio is between \$-0.10 and \$0.10. In these instances we record a profit of \$0.00 and skip to the next day.

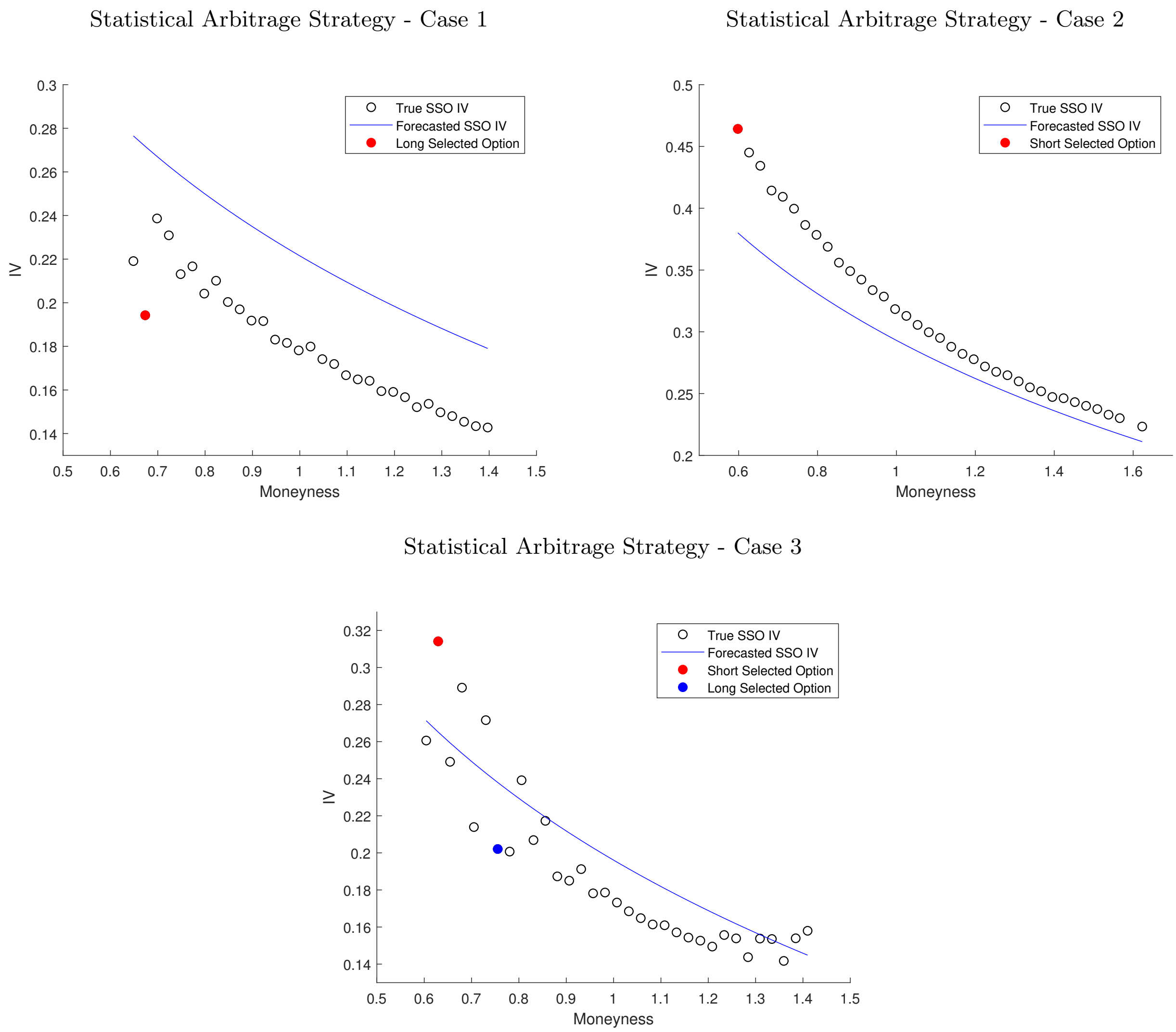


Figure 4: The three possible cases of our statistical arbitrage strategy are shown in these graphs. Both the strings and dates are arbitrarily chosen, purely for presentational purpose. The forecasted values are modeled using the Leung & Sircar method.

All of our models, except for the random walk model, are tested in this economical setting on all four of our LETFs and their options. The random walk model is not included, because its forecasted IV and current IV are the same, whereas our strategy is built on the difference between these two. We present the cumulative trading profits of our models over time in figure 5. We see that the statistical arbitrage strategy is able to achieve gradual profits with a slight jump every now and then for all of the models. Unfortunately, we have to conclude that for none of the LETF options the Leung & Sircar method is able to achieve more trading profits than the other models. It is interesting to see that all four models achieve relatively the same trading profits on the LETF options with a negative leverage factor (SDS and SPXU), whereas the direct estimation approach and the Bernales & Guidolin model outperform the other models for the LETF options with a positive leverage factor.

In table 8 we present the summary statistics of our achieved trading profits from the statistical arbitrage strategy. The Sharpe ratio is defined as the average return earned in excess of the riskfree rate measured per unit of volatility. Therefore we calculate it by first subtracting the riskfree rate from the mean profit and then dividing this by its standard deviation. Furthermore, we count the number of times in which we took a net long position ($V_t \geq \$1.00$), a net short position ($V_t \leq \$-1.00$) or no position (($\$-1.00 < V_t < \$1.00$). Lastly we present the percentage of times that our portfolio selection delivered a positive daily gain. It has to be noted that our trading strategy is delivering a much lower mean profit on the SDS option set than on any of the other LETF options. As can be seen for all LETF options the trading profits and Sharpe ratios for the direct estimation and Bernales & Guidolin method are higher than for the Leung & Sircar method. Even though the results of the Bernales & Guidolin method and direct estimation are close, it is worth mentioning that the Bernales & Guidolin method is beating all other models, both in mean profit and Sharpe ratio, on the SDS and UPRO options. This suggests that the Bernales & Guidolin method proofs, at least somewhat, useful when used in practice. The overall results are as expected, because the performance of the models on the MCPDC is highly correlated to the profits of these kind of trading strategies. Therefore we expected the results to be in line with the MCPDC results. Lastly we would like to point out that even though the trading profits seem to be very promising it is most likely that they turn into losses as soon as we start accounting for trading fees, as can often be seen in other literature.

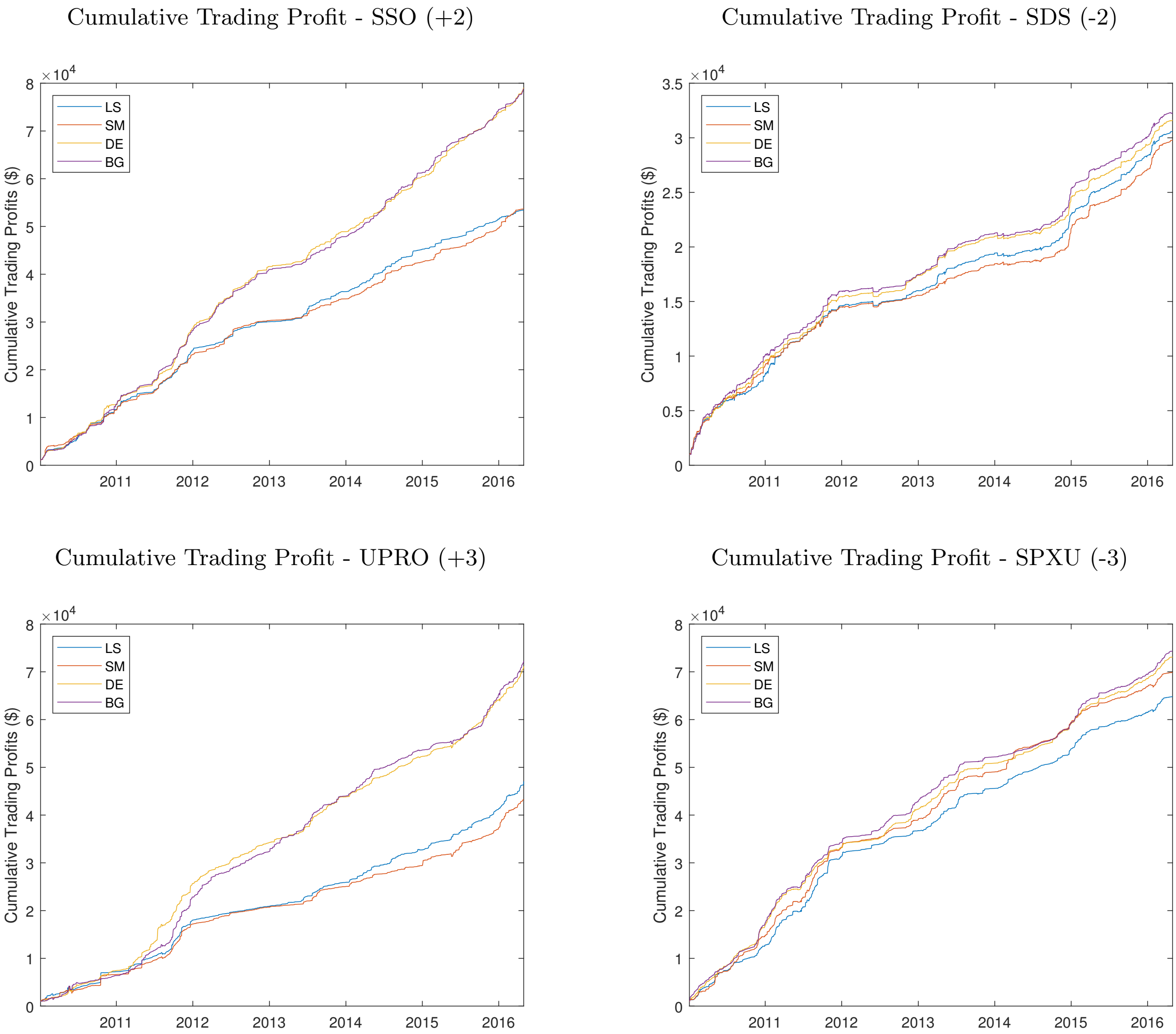


Figure 5: The cumulative trading profits from the statistical arbitrage strategy that are realized by each model are shown in these graphs. The used models are the Leung & Sircar method (LS), direct estimation on LETF options (DE, benchmark 1), the Bernales & Guidolin method (BG, benchmark 2) and the Strawman (SM, benchmark 4). The random walk model does not have any results, because its forecasted IV and current IV are the same, whereas our strategy is built on the difference between these two.

Table 8: Summary Statistics of Statistical Arbitrage Strategy

	LS	**DE**	**BG**	**SM**
SSO (+2)				
Mean Profit (%)	3.31	4.91	4.90	3.32
Std. Dev. Profit (%)	8.48	10.37	10.57	8.82
Sharpe Ratio (%)	37.98	46.47	45.45	36.57
Net Long Positions	531	516	500	519
Net Short Positions	1,023	1,016	1,039	1,025
No Position Taken	33	55	48	43
Positive Gains (%)	81.55	84.01	84.19	78.72
SDS (-2)				
Mean Profit (%)	1.86	1.93	1.97	1.81
Std. Dev. Profit (%)	6.19	5.90	5.98	5.84
Sharpe Ratio (%)	28.58	31.10	31.38	29.40
Net Long Positions	112	105	111	203
Net Short Positions	1,370	1,389	1,381	1,296
No Position Taken	105	93	95	88
Positive Gains (%)	70.91	72.73	72.61	71.16
UPRO (+3)				
Mean Profit (%)	2.90	4.42	4.49	2.66
Std. Dev. Profit (%)	9.25	11.39	10.82	9.46
Sharpe Ratio (%)	30.34	37.96	40.63	27.12
Net Long Positions	465	767	747	432
Net Short Positions	1,100	782	793	1,140
No Position Taken	20	36	45	13
Positive Gains (%)	79.85	82.49	80.16	77.14
SPXU (-3)				
Mean Profit (%)	4.02	4.54	4.62	4.34
Std. Dev. Profit (%)	9.75	8.98	9.48	10.50
Sharpe Ratio (%)	40.24	49.51	47.71	40.41
Net Long Positions	98	164	170	170
Net Short Positions	1,410	1,339	1,329	1,355
No Position Taken	78	83	87	61
Positive Gains (%)	74.56	76.64	76.07	75.44

Notes: This table contains the summary statistics of the achieved trading profits from the statistical arbitrage strategy. The used models are the Leung & Sircar method (LS), direct estimation on LETF options (DE, benchmark 1), the Bernales & Guidolin method (BG, benchmark 2) and the Strawman (SM, benchmark 4). The random walk model is not included, because its forecasted IV and current IV are the same, whereas our strategy is built on the difference between these two.

5. CONCLUSIONS

In this section we summarize the conclusions that can be made from our research, in which we attempt to model and forecast the IVS of LETF options by utilizing the information-rich and highly liquid ETF option data. The main focus here lies on assessing the performance of the method proposed by Leung & Sircar (2015), as their own empirical research is limited. Furthermore we adopt the method proposed by Bernales & Guidolin (2014) on the combination of LETF and ETF options, as this has not been done before in existing literature. We conduct our research on options of (L)ETFs that track the S&P 500, consisting of the regular SPDR S&P500 (SPY) ETF and the following four LETFs: ProShares Ultra S&P500 (SSO), ProShares UltraShort S&P500 (SDS), ProShares UltraPro S&P500 (UPRO) and ProShares UltraPro Short S&P500 (SPXU), with betas of +2, -2, +3 and -3 respectively. After filtering and cleaning the dataset we are left with 1715 trading days, stretching from July 6 2009 to April 29 2016. By means of a statistical and economical evaluation we analyzed the performance of the Leung & Sircar method (LS), direct estimation on LETF options (DE, benchmark 1), the Bernales & Guidolin method (BG, benchmark 2), the random walk (RW, benchmark 3) and the Strawman (SM, benchmark 4).

For our in-sample evaluation we measured the performance of the Leung & Sircar method and the direct estimation approach by means of the RMSE and MAE. The results show that the Leung & Sircar method performs much worse than the direct estimation approach. For both models we use the same deterministic function to model the IVS, however the difference lies in the estimation of its coefficients. Since we are using the ETF option data as a proxy to capture the information of the LETF IVS it would be hard to beat a method that directly models the IVS of LETF options directly from its own data. However, when we analyze the in-sample performance on different segments of the IVS we find that the Leung & Sircar method performs better than the direct estimation approach for several segments for SSO and SDS options.

Next we performed an out-of-sample evaluation by using a recursive moving-window approach to find the one-day-ahead forecasts. The out-of-sample performance is measured by the RMSPE, MAPE and MCPDC. In this dynamic setting we find that the performance of the Leung & Sircar method is somewhat more favorable. Particularly for the set of SDS options the Leung & Sircar method outperforms the other benchmark models in some segments. Moreover, the performance of the Leung & Sircar method on the long-term SSO options is better than the performance of the other benchmark models. Unfortunately its performance on the UPRO and SPXU options is still rather poor. In most of these cases the direct estimation method has the best performance, whereas in some cases the Bernales & Guidolin method has a better performance.

To analyze whether the forecast performance of our models can be utilized we evaluate them in economical sense by using a simple statistical arbitrage strategy. Using this strategy we construct a portfolio that is delta-hedged by its own underlying asset. Unfortunately, we have to conclude that for none of the LETF options the Leung & Sircar method is able to achieve more trading profits than the other models. It is interesting to see that all four models achieve relatively the same trading profits on the LETF options with a negative leverage factor (SDS and SPXU), whereas the direct estimation approach and the Bernales & Guidolin model outperform the other models for the LETF options with a positive leverage factor. Furthermore it has to be noted that our strategy is delivering a significantly lower mean profit on the SDS option set than on any of the other LETF options. Lastly it is worth mentioning that the Bernales & Guidolin model is beating all other models, both in mean profit and Sharpe ratio, on both the SDS and UPRO options.

Overall this research shows that the Leung & Sircar method works in practice, however we were unable to find any general occasions in which the Leung & Sircar method should be preferred over the use of the direct estimation approach. The adoption of the Bernales & Guidolin method on the combination of ETF and LETF options has proven more useful, as it outperforms the direct estimation approach is several cases, both in statistical and economical sense. However, we have to note that the results are rather close and therefore we can not make any definite conclusions on this matter.

6. RECOMMENDATIONS

Naturally, this research evokes several extensions. Therefore we also provide concrete recommendations for further research, following from both the narrow scope of this research and possible improvements on the methods used.

- **Leung & Sircar Method** Even though our results were not particularly promising for the Leung & Sircar method further research should be conducted on the usability of this method on different datasets, e.g. on very illiquid LETF options. We suggest this idea, because in this case the direct estimation approach will not be able to fully capture the dynamics of the IVS of LETF options due to lack of data points. Overall it would be interesting to be able to compare how the Leung & Sircar method performs on LETF options with different ranges of liquidity. A possible improvement would be to estimate the model on either a set of call options or a set of put options, and not on both at the same time. Moreover the used deterministic function is only able to reproduce the implied volatility skew (asymmetric smile), and therefore it is unable to completely capture the volatility smile. This is mostly a problem when estimating the model on a broad range of moneyness coordinates. A solution would be to estimate the model on a dataset with a smaller subset of moneyness coordinates. To elaborate further on this, our dataset has a broad subsection of moneyness coordinates, whereas in other literature on the IV of options it is common to restrict the dataset to a small subset of moneyness coordinates (e.g. a moneyness between 0.9 and 1.1). Another solution would be, all though being more burdensome, to rework the derivations of Leung & Sircar (2015) in an attempt to come up with a different first order approximation for the IV of LETF options. Lastly, further insight could be gained from investigating why the Leung & Sircar method has a particular bad performance on the SSO and UPRO options (the LETFs with a positive leverage factor).

- **Bernales & Guidolin Method** Our research shows that the adoption of the Bernales & Guidolin method on the combination of ETF and LETF options has proven useful in some occasions. Since our set of LETF options are still relatively liquid, it is definitely interesting to conduct further research on the performance of this method more illiquid LETF options. Furthermore, we adopted this method as a benchmark model for the Leung & Sircar method in our research. To allow for comparison we used the same deterministic function as was used by the Leung & Sircar method. Therefore it is advisable to test whether the use of the same deterministic function as given in Bernales & Guidolin (2014) is able to provide a better performance. Moreover, a different moneyness measure could be tested, such as the forward moneyness or the usage of the delta as the moneyness measure as proposed by van der Wel et al. (2016). All together, due to its promising results and the fact that there is no literature available on the usage of the Bernales & Guidolin method on the combination of LETF and ETF options we suggest that more empirical research should be conducted on this matter.

- **Statistical Arbitrage Strategy** First we would like to point out that we estimated our models on a dataset consisting of both call and put options, whereas our strategy has been restricted to only using call options. Therefore we recommend to reestimate the models on a restricted dataset consisting of call options only and then applying the same trading strategy. Most likely this would increase the trading profits, as this should lead to more accurate forecasts of the IVS. Furthermore it is advisable to implement an additional strategy built on statistical arbitrage, e.g. by constructing straddle portfolios as used by Bernales & Guidolin (2014), to analyze whether the results are consistent. Another possibility would be to extend the used strategy by accounting for other option price sensitivities, by gamma, theta or charm-adjusted delta hedging. We would also like to point out that it would be useful to have a simple benchmark model to evaluate our models against in this economical setting, e.g. a benchmark model based on the historical average of the IVS. Due to time constraints of this master thesis we leave this suggestion for further research. Lastly we recommend to do the analysis again, but this time also take trading fees and the effective spread into consideration. We make this recommendation, because in most other literature it turns out that trading profits will disappear when accounting for trading fees and we do not expect it to be different for this research.

References

Marco Avellaneda and Stanley Zhang. Path-dependence of leveraged etf returns. *SIAM Journal on Financial Mathematics*, 1(1):586–603, 2010.

Alejandro Bernales and Massimo Guidolin. Can we forecast the implied volatility surface dynamics of equity options? predictability and economic value tests. *Journal of Banking & Finance*, 46:326–342, 2014.

Fischer Black and Myron Scholes. The pricing of options and corporate liabilities. *Journal of political economy*, 81(3):637–654, 1973.

Minder Cheng and Ananth Madhavan. The dynamics of leveraged and inverse exchange-traded funds. 2010.

Rama Cont, José Da Fonseca, et al. Dynamics of implied volatility surfaces. *Quantitative finance*, 2(1):45–60, 2002.

John C Cox, Stephen A Ross, and Mark Rubinstein. Option pricing: A simplified approach. *Journal of financial Economics*, 7(3):229–263, 1979.

Bernard Dumas, Jeff Fleming, and Robert E Whaley. Implied volatility functions: Empirical tests. *The Journal of Finance*, 53(6):2059–2106, 1998.

Matthias R Fengler, Wolfgang K Härdle, and Enno Mammen. A semiparametric factor model for implied volatility surface dynamics. *Journal of Financial Econometrics*, 5(2):189–218, 2007.

Jean-Pierre Fouque, George Papanicolaou, Ronnie Sircar, and Knut Sølna. *Multiscale stochastic volatility for equity, interest rate, and credit derivatives.* Cambridge University Press, 2011.

Silvia Goncalves and Massimo Guidolin. Predictable dynamics in the s&p 500 index options implied volatility surface. *The Journal of Business*, 79(3):1591–1635, 2006.

Wolfgang K Härdle, Sergey Nasekin, and Zhiwu Hong. Leveraged etf options implied volatility paradox: a statistical study. 2016.

Ludger Hentschel. Errors in implied volatility estimation. *Journal of Financial and Quantitative analysis*, 38(4): 779–810, 2003.

Tim Leung and Marco Santoli. *Leveraged Exchange-Traded Funds: Price Dynamics and Options Valuation.* Springer, 2016.

Tim Leung and Ronnie Sircar. Implied volatility of leveraged etf options. *Applied Mathematical Finance*, 22(2): 162–188, 2015.

Michel van der Wel, Sait R Ozturk, and Dick van Dijk. Dynamic factor models for the volatility surface. In *Dynamic Factor Models*, pages 127–174. Emerald Group Publishing Limited, 2016.

APPENDIX A. DERIVATIONS IMPLIED VOLATILITY SCALING

This appendix summarizes the idea behind implied volatility scaling, as presented in Leung & Sircar (2015). First they show that theoretical call prices on the LETF L can be expressed in terms of the Black-Scholes formula, with its volatility scaled by the absolute value of the leverage ratio β, as given by:

$$C_{\mathrm{BS},\beta}(t, L; K, T) = C_{\mathrm{BS}}(t, L; K, T, r, c, |\beta|\sigma) \tag{20}$$

where $C_{\mathrm{BS}}(t, L; K, T, r, c, \sigma)$ is the standard Black-Scholes formula for a call option with strike price K, expiration date T, interest rate r, dividend rate c and volatility parameter σ. The implied volatility of a LETF option is given in terms of the inverse of the Black-Scholes formula:

$$\widetilde{IV}_{\mathrm{LETF}} = (C_{\mathrm{BS},\beta})^{-1}(C_{\mathrm{obs}}) = \frac{1}{|\beta|} C_{BS}^{-1}(C_{\mathrm{obs}}) \tag{21}$$

where C_{obs} is the observed option price. They note that they normalize by $|\beta|^{-1}$ in their definition of the implied volatility so that they remain on the same scale. Hypothetically, if the Black-Scholes model price and the observed option price are the same, then it follows from equation 20 that the implied volatility reduces to:

$$\widetilde{IV}_{\mathrm{LETF}} = \frac{1}{|\beta|} C_{\mathrm{BS}}^{-1}(C_{\mathrm{BS}}(t, L; K, T, r, c, |\beta|\sigma)) = \frac{1}{|\beta|} |\beta|\sigma = \sigma \tag{22}$$

Thus, if the market prices follow the Black-Scholes model, all calls written on ETFs and corresponding LETFs must have an identical implied volatility (same conclusions goes for put options).

APPENDIX B. DERIVATIONS MONEYNESS SCALING

This appendix summarizes the derivations for the moneyness scaling approach as presented in Leung & Sircar (2015). First they write the log LETF price, in a general stochastic volatility model, as:

$$\log\left(\frac{L_T}{L_0}\right) = \beta\log\left(\frac{X_T}{X_0}\right) - (r(\beta-1)+c)T - \frac{\beta(\beta-1)}{2}\int_0^T \sigma_t^2 dt \tag{23}$$

where $(\sigma_t)_{t\geq 0}$ is the stochastic volatility process. Furthermore, $\log\left(\frac{L_T}{L_0}\right)$ and $\log\left(\frac{X_T}{X_0}\right)$ are the log-moneyness of the terminal LETF and ETF values respectively. Next they introduce the condition that the terminal log-moneyness $\log\left(\frac{X_T}{X_0}\right)$ is equal to some constant LM_{ETF}. Then equation 23 changes towards:

$$LM_{\text{LETF}} = \beta LM_{\text{ETF}} - (r(\beta-1)+c)T - \frac{\beta(\beta-1)}{2} I\!E\left\{\int_0^T \sigma_t^2 dt \middle| \log\left(\frac{X_T}{X_0}\right) = LM_{\text{ETF}}\right\} \tag{24}$$

In order to obtain a more explicit relationship they consider the case with a constant σ, as is also done in the Black-Scholes model. Then equation 24 reduces to:

$$LM_{\text{LETF}} = \beta LM_{\text{ETF}} - (r(\beta-1)+c)T - \frac{\beta(\beta-1)}{2}\sigma^2 T \tag{25}$$

In the same way they derive an equation to link the log-moneyness pair of two LETFs with different leverage ratios, which is given by:

$$LM_{\text{LETF}_1} = \frac{\beta_1}{\beta_2}\left[LM_{\text{LETF}_2} + (r(\beta_2-1)+c_2)T + \frac{\beta_2(\beta_2-1)}{2}\bar{\sigma}^2 T\right] - (r(\beta_1-1)+c_1)T - \frac{\beta_1(\beta_1-1)}{2}\bar{\sigma}^2 T \tag{26}$$

APPENDIX C. DERIVATIONS LEUNG & SIRCAR METHOD

This appendix summarizes the derivations that are needed for the Leung & Sircar method, as presented in Leung & Sircar (2015). The class of option pricing models that is used for these derivations comes from Fouque et al. (2011), and thus further technical details and derivations can be found there. First of all they present the following system of SDEs, under a risk neutral measure IP, with reference index X, the fast volatility factor Y and the slow volatility factor Z:

$$\begin{aligned} dX_t &= rX_t dt + f(Y_t, Z_t)X_t dW_t^{(0)} \\ dY_t &= \left(\frac{1}{\epsilon}\alpha(Y_t) - \frac{1}{\sqrt{\epsilon}}\eta(Y_t)\Lambda_1(Y_t, Z_t)\right)dt + \frac{1}{\sqrt{\epsilon}}\eta(Y_t)dW_t^{(1)} \\ dZ_t &= \left(\delta\theta(Z_t) - \sqrt{\delta}g(Z_t)\Lambda_2(Y_t, Z_t)\right)dt + \sqrt{\delta}g(Z_t)dW_t^{(2)} \end{aligned} \tag{27}$$

where the standard IP-Brownian motions $(W^{(0)}, W^{(1)}, W^{(2)})$ are correlated as follows:

$$\langle W^{(0)}, W^{(0)}\rangle_t = \rho_1 dt \qquad \langle W^{(0)}, W^{(2)}\rangle_t = \rho_2 dt \qquad \langle W^{(1)}, W^{(2)}\rangle_t = \rho_{12} dt$$

with the usual conditions to ensure positive definiteness of the associated covariance matrix. In the system of SDEs $\Lambda_1(Y_t, Z_t)$ and $\Lambda_2(Y_t, Z_t)$ are the associated market prices of volatility risk. The fast volatility factor Y is characterized by the small positive parameter ϵ, whereas the slow volatility factor Z is characterized by the small positive parameter δ. The other coefficients are not of interest for the rest of the analysis.

Next they use the perturbation theory to approximate the no arbitrage price of an option. The full steps taken for this can be found in chapters 4-5 of Fouque et al. (2011), but we will shortly summarize the path that is followed. They first state that by $\langle\cdot\rangle$ they denote averaging with respect to the invariant distribution Φ. Next they define the following four stochastic volatility group parameters:

1. The corrected volatility parameter $\sigma(z)$, which is composed of the averaged effective volatility $\bar{\sigma}^2(z)$ adjusted by a small term from the market price of volatility risk from the fast factor Y ($V_2(z)$). Here $V_2(z)$ is of order $\sqrt{\epsilon}$. Both the definitions of $\bar{\sigma}^2(z)$ and $V_2(z)$ are irrelevant for the rest of the analysis. For those that are interested, these derivations can be found in chapters 4-5 of Fouque et al. (2011).

$$\sigma(z) = \sqrt{\bar{\sigma}^2(z) + 2V_2(z)}$$

2. The group parameter $V_3(z)$ is of order $\sqrt{\epsilon}$ and contains the correlation ρ_1 between shocks to the fast factor Y and the index X.

$$V_3(z) = -\frac{\rho_1\sqrt{\epsilon}}{2}\langle\eta f\frac{\delta\phi}{\delta y}\rangle$$

3. The group parameter V_0 is of order $\sqrt{\delta}$ and contains the market price of volatility risk from the slow factor Z.

$$V_0(z) = -\frac{g(z)\sqrt{\delta}}{2}\langle\Delta_2\rangle\bar{\sigma}'(z)$$

4. The group parameter V_1 is of order $\sqrt{\delta}$ and contains the correlation ρ_2 between shocks to the slow factor Z and the index X.

$$V_1(z) = -\frac{\rho_2\; g(z)\sqrt{\delta}}{2}\langle f\rangle\bar{\sigma}'(z)$$

All of the above definitions are then used to derive a formula to approximate an option price. This formula has the following representation:

$$P = P_{BS} + \left\{ \tau V_0 + \tau V_1 (x \frac{\delta}{\delta x} + \frac{V_3}{\sigma}(x \frac{\delta}{\delta x}) \right\} \frac{\delta P_{BS}}{\delta \sigma} \tag{28}$$

This price approximation can then be used to derive the first order approximation of the implied volatility, which at one point in time is presented by:

$$IV_{\text{ETF}} = b_1 + b_2 \tau + a_1 \frac{LM_{\text{ETF}}}{\tau} + a_2 LM_{\text{ETF}} + \epsilon \tag{29}$$

where the parameters (b_1, b_2, a_1, a_2) are defined in terms of the stochastic volatility group parameters (σ, V_0, V_1, V_3) by the following formulas:

$$\begin{aligned} b_1 &= \sigma + \frac{V_3}{2\sigma}\left(1 - \frac{2r}{\sigma^2}\right) & a_1 &= \frac{V_3}{\sigma^3} \\ b_2 &= V_0 + \frac{V_1}{2}\left(1 - \frac{2r}{\sigma^2}\right) & a_2 &= \frac{V_1}{\sigma^2} \end{aligned} \tag{30}$$

where (a_1, a_2) are the the skew slope parameters, whereas (b_1, b_2) are the level parameters. Furthermore r is the yearly risk-free rate. For the complete derivation of these formulas we refer to chapters 4-5 of Fouque et al. (2011).

Leung & Sircar (2015) extended these derivations for the IV of LETF options. First of all they added the following SDE for the LETF L into the system of SDEs from equation 27:

$$dL_t = (r - c)L_t dt + \beta f(Y_t, Z_t) L_t dW_t^{(0)} \tag{31}$$

By following the same steps as for the regular ETF options they derived an approximation for the LETF option price, similar to equation 28, however this time including the influence of the β of its associated LETF by means of:

$$P_\beta = P_{BS} + \left\{ \tau|\beta|V_0 + \tau\beta|\beta|V_1 (x \frac{\delta}{\delta x} + \frac{\beta^3 V_3}{|\beta|\sigma}(x \frac{\delta}{\delta x}) \right\} \frac{\delta P_{BS}}{\delta \sigma} \tag{32}$$

Equivalently this price approximation is used to derive the first order approximation for the IV of a LETF option at one point in time is given by:

$$\widetilde{IV}_{\text{LETF}} = b_{1,\beta} + b_{2,\beta}\tau + a_{1,\beta} \frac{LM_{\text{LETF}}}{\tau} + a_{2,\beta} LM_{\text{LETF}} + \epsilon \tag{33}$$

Note that the log-moneyness in this approximation is the unscaled variant. The parameters $(b_{1,\beta}, b_{2,\beta}, a_{1,\beta}, a_{2,\beta})$ are defined in terms of the same stochastic volatility group parameters (σ, V_0, V_1, V_3) as its unleveraged variant. Leung & Sircar (2015) derived the following formulas to show how the β interacts with the stochastic volatility group parameters:

$$\begin{aligned} b_{1,\beta} &= \sigma + \frac{\beta V_3}{2\sigma}\left(1 - \frac{2r}{\beta^2\sigma^2}\right) & a_{1,\beta} &= \frac{V_3}{\beta\sigma^3} \\ b_{2,\beta} &= V_0 + \frac{\beta V_1}{2}\left(1 - \frac{2r}{\beta^2\sigma^2}\right) & a_{2,\beta} &= \frac{V_1}{\beta\sigma^2} \end{aligned} \tag{34}$$

APPENDIX D. SPLIT HISTORY OF THE (L)ETFS

In this appendix we present the split history of our LETF shares. It is common practice for LETFs to (reverse) split its number of shares when the price has reached a certain level. This is done to ensure that its price will remain in a convenient trading range. In table 9 the split history of the four LETFs can be seen. A split of 2:1 means that on that specific date every share holder receives double the number of shares, causing a 50% decrease of its price. Once such a split is announced, the corresponding options are altered in such a way that it has neither a positive or negative effect for its holders.

Table 9: Split history of LETF shares

SSO (+2)	**SDS (-2)**	**UPRO (+3)**	**SPXU (-3)**
May 20, 2015 - 2:1	Oct 5, 2012 - 1:4	Feb 25, 2011 - 3:1	May 11, 2012 - 1:5
		Jun 10, 2013 - 2:1	Jan 24, 2014 - 1:4
		May 20, 2015 - 2:1	

Notes: The table shows the different (reverse) splits that occurred for the shares of the LETFs. A split of 2:1 means that on that specific date every share holder receives double the number of shares, causing a 50% decrease in its price.

APPENDIX E. OPTION TRADING VOLUME AND AVAILABILITY

In this appendix we present how both the daily aggregated trading volume of options on (L)ETFs and the number of available contracts have changed over the timespan of our dataset. The average daily aggregated trading volume of options on (L)ETFs is shown in table 10, whereas the number of available option contracts on (L)ETFs is shown in table 6. For the trading volume the most important observations here are that the trading volume on SPY, UPRO and SPXU options grew rapidly until 2011, whereafter it stayed relatively on the same level. For both SSO and SDS options the trading volume grew until 2011 and declined thereafter. For the number of available contracts we conclude that the number of available option contracts for the SPY ETF has grown steadily over time, from 2101 on July 6th 2009 to 3731 on April 29th 2016. Furthermore, the number of available option contracts for SSO and UPRO has also grown, whereas the number of available option contracts for SDS and SPXU maintained relatively on the same level. Interestingly enough, despite the growth of available option contracts we find that the aggregated trading volume on each trading day did not grow over this period for any of the (L)ETF options.

Table 10: The average daily aggregated trading volume of options on (L)ETFs per year.

	SPY	**SSO (+2)**	**SDS (-2)**	**UPRO (+3)**	**SPXU (-3)**
2009	1,201,800	15,021	29,728	509	1,659
2010	1,528,400	17,990	38,331	496	958
2011	2,585,700	19,960	53,048	1,482	3,592
2012	2,063,400	13,559	36,864	1,072	2,798
2013	2,115,200	14,682	33,367	1,495	3,585
2014	2,148,000	8,831	23,367	1,333	1,241
2015	2,317,800	3,832	16,303	1,349	1,859
2016	2,405,000	4,063	17,248	1,360	2,031

Notes: This table contains the average of the daily aggregated trading volume of options on (L)ETFs, shown for each year. The data is based on 1715 trading days spanning July 6th 2009 to April 29th 2016.

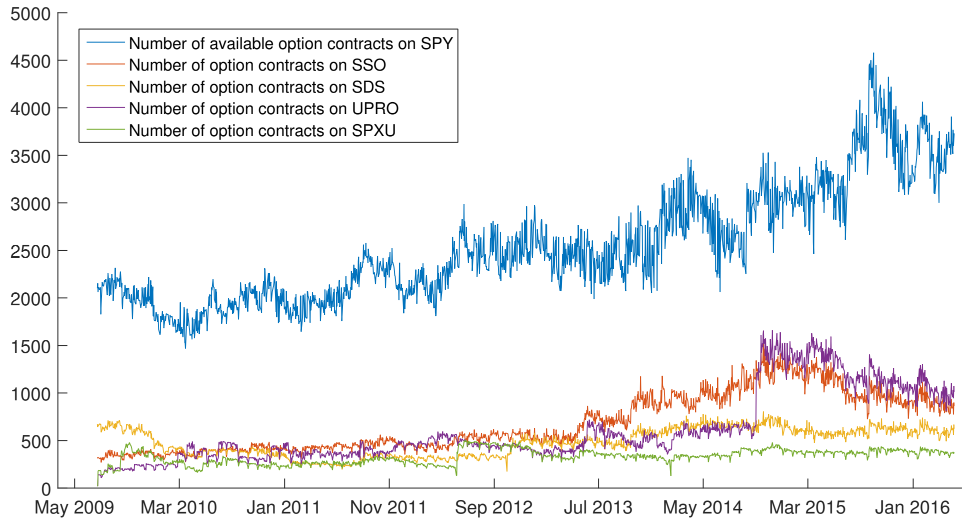
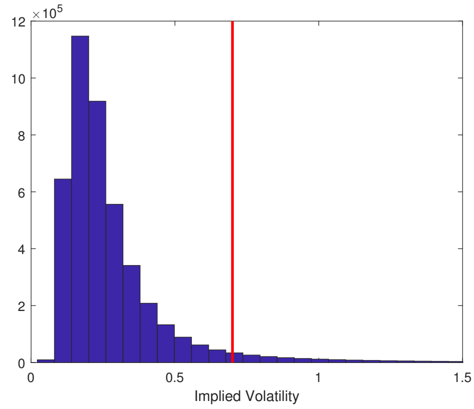


Figure 6: The daily number of available option contracts on each ETF over time.

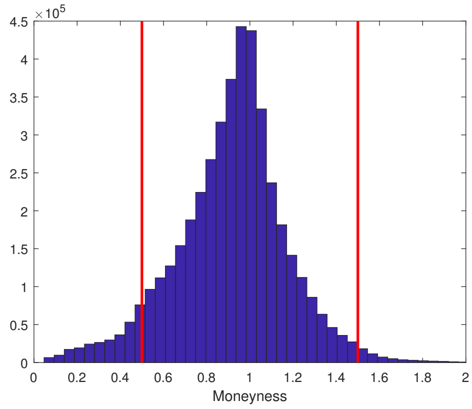
APPENDIX F. HISTOGRAMS OF IV DATA

In this appendix we provide empirical verification of the filters that we use on the (scaled) implied volatility and the (scaled) moneyness. Due to the nature of LETFs we had to adjust the filters that are used in other literature accordingly. Therefore the empirical verification of the used filters on the (scaled) implied volatility and the (scaled) moneyness is shown below in figure 8 by means of the histogram plots.

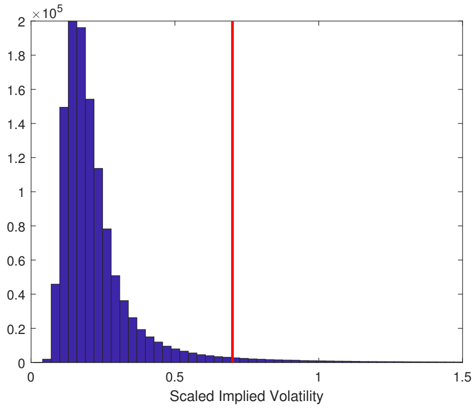
SPY Implied Volatility



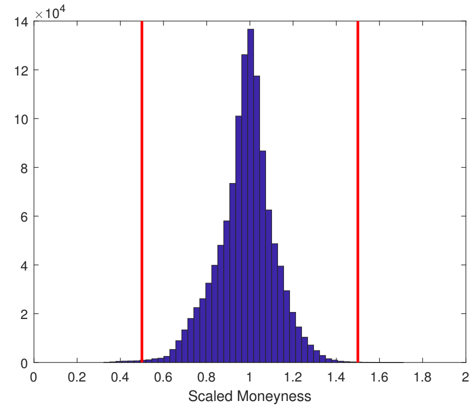
SPY Moneyness



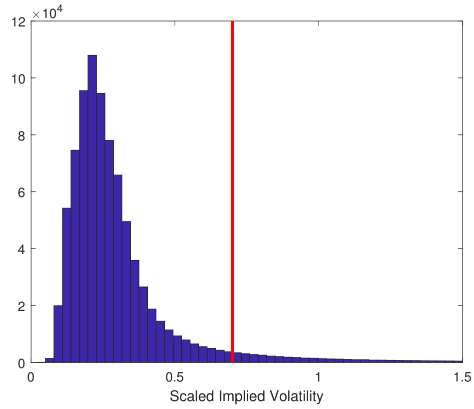
SSO Scaled Implied Volatility



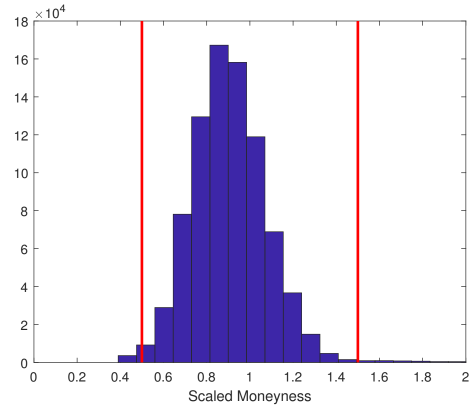
SSO Scaled Moneyness



SDS Scaled Implied Volatility



SDS Scaled Moneyness



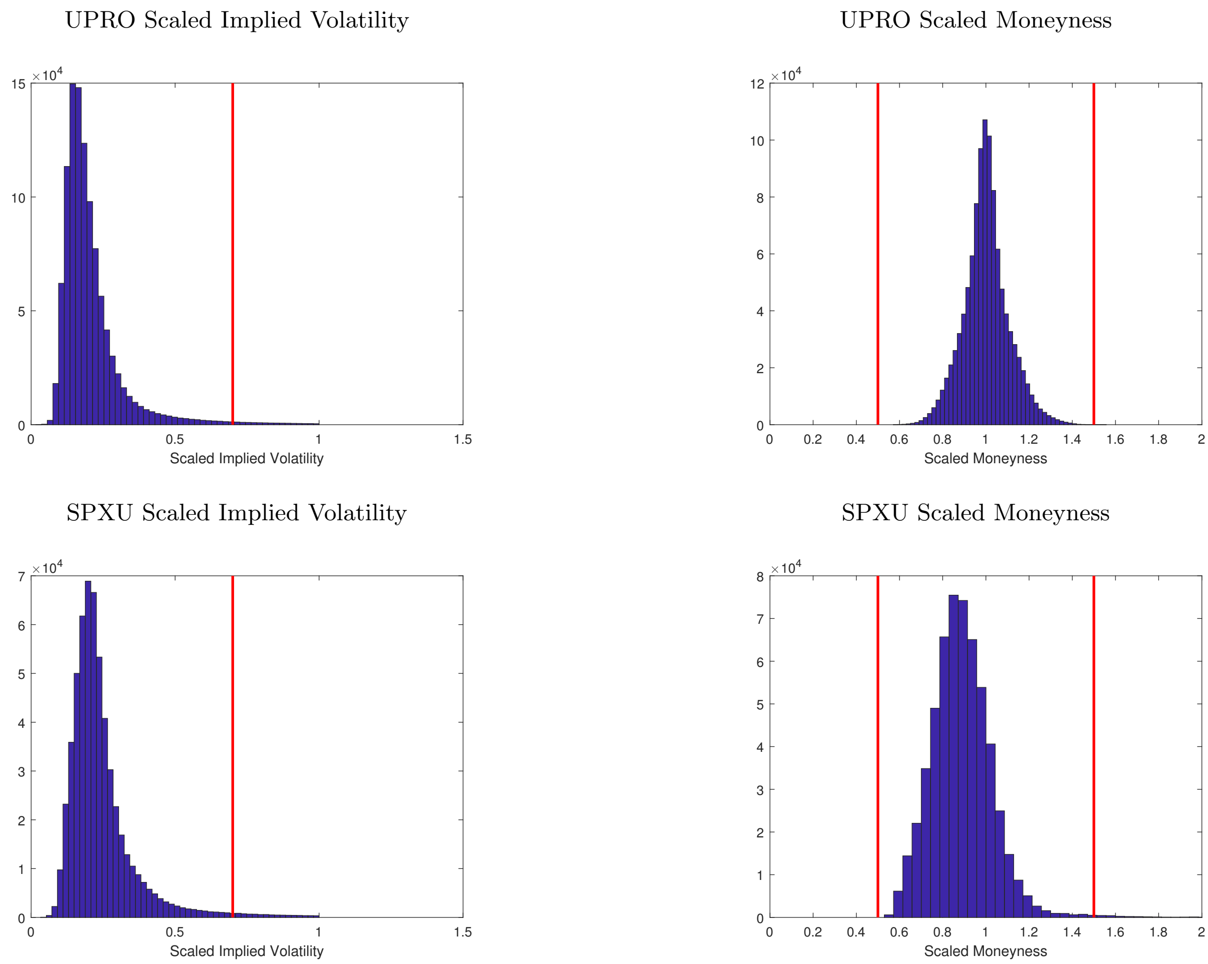


Figure 8: The histograms for the SPY and LETF (scaled) implied volatility and (scaled) moneyness.

APPENDIX G. LOST OBSERVATIONS PER FILTER

In this appendix we presents the number of lost observations per filter, as can be seen in table 11 . Each of these filters is separately applied on all observations per (L)ETF. Therefore, it has to be taken into account that there is overlap between the filters, and thus the total observations lost is not the same as the sum of the lost observations per filter.

Table 11: Lost observations per filter

	SPY	**SSO (+2)**	**SDS (-2)**	**UPRO (+3)**	**SPXU (-3)**
Total Observations	4,360,551	1,169,341	827,823	1,049,086	566,022
Total Lost Observations	1,558,449	352,502	319,452	256,380	166,050
	35.7%	30.1%	38.6%	24.4%	29.3%
Total Observations Left	2,802,102	816,839	508,371	792,706	399,972
	64.3%	69.9%	61.4%	75.6%	70.7%
1) TTM < 10 **& TTM** > 365	941,859	246,913	172,899	176,131	96,782
	21.6%	21.1%	20.9%	16.8%	17.1%
2) Price $<$ **\$**0.05	655,844	66,542	112,327	1,415	9,880
	15.1%	5.7%	13.6%	0.1%	1.7%
3) $\widetilde{IV}_{\mathbf{(L)ETF}} < 0.7$	192,445	25,357	36,787	11,449	8,334
	4.4%	2.2%	4.4%	1.1%	1.5%
4) $0.5 < \widetilde{M}_{\mathbf{(L)ETF}} < 1.5$	327,562	3,288	12,919	45	2,611
	7.6%	0.3%	1.6%	0.0%	0.5%
5) Bid Quote = **\$**0	450,040	143,733	168,120	119,761	78,997
	10.3%	12.3%	20.3%	11.4%	14.0%

Notes: In the first section of the table we show the total number of lost observations, when all filters are applied at the same time. In the second part we apply each of these filters separately on all observations per (L)ETF. The data covers the period between July 6th 2009 to April 29th 2016.

APPENDIX H. DATA DISTRIBUTION OVER DIFFERENT SEGMENTS

In this appendix the distribution of our filtered dataset over the different segments of the IVS is shown, as can be seen below in table H. It can be seen that the far-in-the-money options cover a relatively small part of the dataset. Furthermore the distribution gets less dense as the time to maturity increases.

Table 12: Number of observations over different segments of the IVS.

	Short Term (10-90 days)		**Medium Term (90-180 days)**		**Long Term (180-365 days)**	
	Observations	**(%)**	**Observations**	**(%)**	**Observations**	**(%)**
SPY (+1)						
Total: 2,801,978						
$0.5 \leq M \leq 0.9$	353,369	12.6	343,773	12.3	414,784	14.8
$0.9 \leq M \leq 1$	353,905	12.6	148,606	5.3	221,402	7.9
$1.0 \leq M \leq 1.1$	262,825	9.4	129,690	4.6	178,602	6.4
$1.1 \leq M \leq 1.5$	94,332	3.4	108,013	3.9	192,644	6.9
SSO (+2)						
Total: 820,457						
$0.5 \leq \widetilde{M} \leq 0.9$	71,145	8.7	60,496	7.4	48,098	5.9
$0.9 \leq \widetilde{M} \leq 1$	196,453	24.0	49,743	6.1	38,747	4.7
$1.0 \leq \widetilde{M} \leq 1.1$	159,643	19.5	50,029	6.1	42,492	5.2
$1.1 \leq \widetilde{M} \leq 1.5$	23,014	2.8	38,902	4.8	38,077	4.6
SDS (-2)						
Total: 512,625						
$0.5 \leq \widetilde{M} \leq 0.9$	93,374	18.4	85,769	16.9	81,563	16.0
$0.9 \leq \widetilde{M} \leq 1$	94,845	18.7	26,139	5.1	21,019	4.1
$1.0 \leq \widetilde{M} \leq 1.1$	45,379	8.9	16,694	3.3	12,867	2.5
$1.1 \leq \widetilde{M} \leq 1.5$	16,470	3.2	7,782	1.5	6,472	1.3
UPRO (+3)						
Total: 793,571						
$0.5 \leq \widetilde{M} \leq 0.9$	49,490	6.2	39,322	5.0	20,402	2.6
$0.9 \leq \widetilde{M} \leq 1$	203,684	25.7	64,988	8.2	41,665	5.3
$1.0 \leq \widetilde{M} \leq 1.1$	169,803	21.4	56,645	7.2	58,184	7.3
$1.1 \leq \widetilde{M} \leq 1.5$	16,833	2.1	33,926	4.3	37,764	4.8
SPXU (-3)						
Total: 401,030						
$0.5 \leq \widetilde{M} \leq 0.9$	62,222	15.6	80,130	20.0	68,984	17.3
$0.9 \leq \widetilde{M} \leq 1$	68,982	17.3	36,495	9.1	23,568	5.9
$1.0 \leq \widetilde{M} \leq 1.1$	26,902	6.7	14,372	3.6	8,192	2.1
$1.1 \leq \widetilde{M} \leq 1.5$	5,411	1.4	2,832	0.7	1,882	0.5

Notes: This table contains the number of observations over the different segments of the IVS. This is based on the filtered dataset that covers the period between July 6th 2009 to April 29th 2016.

APPENDIX I. IN-SAMPLE PERFORMANCE PER SEGMENT - MAE

In this appendix the in-sample performance per segment measured as the Mean Absolute Error (MAE) is shown, as can be found in table 13. The used models are the Leung & Sircar method (LS) and the Direct Estimation on LETF options (DE, benchmark 3). The moneyness of the LETF options is transformed to the same scale as the SPY options.

Table 13: In-Sample Performance per Segment - MAE.

	Short Term (10-90 days)		**Medium Term (90-180 days)**		**Long Term (180-365 days)**	
	LS	**DE**	**LS**	**DE**	**LS**	**DE**
SSO (+2)						
$0.5 \leq \widetilde{M} \leq 0.9$	0.046	0.035	0.023	0.022	0.022	0.023
$0.9 \leq \widetilde{M} \leq 1$	0.040	0.035	0.020	0.015	0.014	0.017
$1.0 \leq \widetilde{M} \leq 1.1$	0.051	0.040	0.025	0.017	0.015	0.016
$1.1 \leq \widetilde{M} \leq 1.5$	0.068	0.054	0.031	0.024	0.019	0.022
SDS (-2)						
$0.5 \leq \widetilde{M} \leq 0.9$	0.126	0.085	0.048	0.040	0.031	0.029
$0.9 \leq \widetilde{M} \leq 1$	0.049	0.054	0.025	0.029	0.016	0.019
$1.0 \leq \widetilde{M} \leq 1.1$	0.036	0.040	0.019	0.021	0.014	0.016
$1.1 \leq \widetilde{M} \leq 1.5$	0.040	0.036	0.016	0.015	0.016	0.013
UPRO (+3)						
$0.5 \leq \widetilde{M} \leq 0.9$	0.038	0.025	0.019	0.016	0.018	0.015
$0.9 \leq \widetilde{M} \leq 1$	0.042	0.019	0.022	0.010	0.014	0.011
$1.0 \leq \widetilde{M} \leq 1.1$	0.051	0.021	0.025	0.011	0.015	0.011
$1.1 \leq \widetilde{M} \leq 1.5$	0.056	0.030	0.029	0.015	0.017	0.013
SPXU (-3)						
$0.5 \leq \widetilde{M} \leq 0.9$	0.079	0.056	0.038	0.027	0.025	0.021
$0.9 \leq \widetilde{M} \leq 1$	0.049	0.038	0.029	0.020	0.019	0.016
$1.0 \leq \widetilde{M} \leq 1.1$	0.039	0.029	0.024	0.016	0.018	0.014
$1.1 \leq \widetilde{M} \leq 1.5$	0.034	0.027	0.019	0.013	0.016	0.013

Notes: This table contains the in-sample fit per segment as measured in the Mean Absolute Error (MAE). The used models are the Leung & Sircar method (LS) and the Direct Estimation on LETF options (DE, benchmark 1). The moneyness of the LETF options is transformed to the same scale as the SPY options.

APPENDIX J. OUT-OF-SAMPLE PERFORMANCE PER SEGMENT - MAPE

In this appendix the out-of-sample performance per segment measured as the Mean Absolute Prediction Error (MAPE) is shown, as can be seen in table 14. The used models are the Leung & Sircar method (LS), direct estimation on LETF options (DE, benchmark 1), the Bernales & Guidolin method (BG, benchmark 2), the Random Walk (RW, benchmark 3) and the Strawman (SM, benchmark 4). The moneyness of the LETF options is transformed to the same scale as the SPY options. For each segment we emboldened the best performing model. It is important to note that we neglected the results of the random walk. We used a six month rolling window of data to find the one-day-ahead predictions.

Table 14: Out-of-Sample Forecast Performance per Segment - MAPE.

	Short Term (10-90 days)					**Medium Term (90-180 days)**					**Long Term (180-365 days)**				
	LS	**DE**	**BG**	**RW**	**SM**	**LS**	**DE**	**BG**	**RW**	**SM**	**LS**	**DE**	**BG**	**RW**	**SM**
SSO (+2)															
$0.5 \leq \widetilde{M} \leq 0.9$	0.035	**0.033**	0.033	0.020	0.041	**0.022**	0.022	0.022	0.015	0.023	**0.020**	0.025	0.026	0.013	0.022
$0.9 \leq \widetilde{M} \leq 1.0$	0.040	0.039	**0.038**	0.011	0.040	0.020		0.016	**0.015** 0.007	0.021	**0.014**	0.019	0.020	0.008	0.015
$1.0 \leq \widetilde{M} \leq 1.1$	0.047	0.041	**0.041**	0.011	0.051	0.024	0.017	**0.016**	0.007	0.026	**0.015**	0.018	0.018	0.007	0.015
$1.1 \leq \widetilde{M} \leq 1.5$	**0.056**	0.055	0.056	0.024	0.064	0.028	0.024	**0.023**	0.013	0.031	**0.018**	0.023	0.024	0.010	0.019
SDS (-2)															
$0.5 \leq \widetilde{M} \leq 0.9$	0.101	0.090	**0.089**	0.037	0.106	**0.040**	0.042	0.042	0.017	0.042	0.026	**0.025**	0.026	0.017	0.027
$0.9 \leq \widetilde{M} \leq 1.0$	**0.044**	0.055	0.055	0.011	0.049	**0.024**	0.029	0.029	0.006	0.026	**0.015**	0.016	0.016	0.007	0.016
$1.0 \leq \widetilde{M} \leq 1.1$	**0.036**	0.043	0.042	0.010	0.036	**0.018**	0.022	0.021	0.005	0.019	**0.013**	0.013	0.013	0.006	0.014
$1.1 \leq \widetilde{M} \leq 1.5$	**0.033**	0.034	0.035	0.022	0.038	0.017	0.016	**0.016**	0.011	0.017	0.015	**0.013**	0.014	0.009	0.017
UPRO (+3)															
$0.5 \leq \widetilde{M} \leq 0.9$	0.034	**0.026**	0.027	0.021	0.036	0.019	**0.018**	0.018	0.012	0.019	**0.017**	0.017	0.017	0.008	0.019
$0.9 \leq \widetilde{M} \leq 1.0$	0.041	0.022	**0.021**	0.013	0.042	0.021	**0.011**	0.012	0.007	0.022	0.014	0.013	**0.013**	0.006	0.015
$1.0 \leq \widetilde{M} \leq 1.1$	0.047	**0.023**	0.023	0.013	0.050	0.024	0.012	**0.012**	0.007	0.025	0.015	0.012	**0.012**	0.006	0.015
$1.1 \leq \widetilde{M} \leq 1.5$	0.047	**0.033**	0.034	0.020	0.055	0.027	**0.016**	0.016	0.009	0.029	0.016	0.014	**0.013**	0.006	0.017
SPXU (-3)															
$0.5 \leq \widetilde{M} \leq 0.9$	0.064	0.060	**0.060**	0.030	0.069	0.035	0.029	**0.028**	0.018	0.036	0.023	**0.020**	0.021	0.017	0.024
$0.9 \leq \widetilde{M} \leq 1.0$	0.045	0.039	**0.039**	0.013	0.048	0.027	0.020	**0.020**	0.011	0.028	0.018	**0.015**	0.016	0.012	0.018
$1.0 \leq \widetilde{M} \leq 1.1$	0.038	0.034	**0.033**	0.012	0.039	0.023	0.017	**0.016**	0.010	0.023	0.016	**0.014**	0.015	0.012	0.017
$1.1 \leq \widetilde{M} \leq 1.5$	0.032	0.029	**0.029**	0.023	0.034	0.019	0.015	**0.014**	0.013	0.019	0.015	**0.014**	0.014	0.012	0.016

Notes: This table contains the out-of-sample forecast performance per segment as measured in the Mean Absolute Prediction Error (MAPE). The used models are the Leung & Sircar method (LS), direct estimation on LETF options (DE, benchmark 1), the Bernales & Guidolin method (BG, benchmark 2), the random walk (RW, benchmark 3) and the Strawman (SM, benchmark 4). The moneyness of the LETF options is transformed to the same scale as the SPY options. For each segment we emboldened the best performing model. It is important to note that we neglected the results of the random walk. We used a six month rolling window of data to find the one-day-ahead predictions.