A Financial Modeling Approach to Industry Exchange-Traded Funds Selection

Thomas Conlon[a], John Cotter[a], Illia Kovalenko[a,*], Thierry Post[b]

[a] *Smurfit School of Business, University College Dublin, Blackrock, Co. Dublin, Ireland*
[b] *Graduate School of Business of Nazarbayev University, Astana, Kazakhstan*

Abstract

This study uses a comprehensive financial modeling approach to optimize the selection of equity sector Exchange Traded Funds. The probability distribution of fund returns is estimated using market options prices, the Heston stochastic volatility model, and risk-premium transformation methods. The probability distribution is discretized using Monte-Carlo simulation, and the simulated input samples are winsorized to account for estimation and simulation error. We use a portfolio optimization with stochastic dominance constraints to account for higher moments. The optimal portfolio outperforms the S&P 500 index, a basic sector-momentum strategy, and a range of alternative active strategies out of sample, after accounting for risk and transaction costs. The superior performance is more pronounced during high-volatility states during which option prices are most informative. No significant performance gains are achieved with a Regular Vine copula instead of a traditional Gaussian copula, despite the non-linear dependence that exists between the funds.

Keywords: Sector Exchange Traded Funds, Portfolio optimization, Option-implied distribution, Copulas, Stochastic dominance

*Corresponding author
Email addresses: illia.kovalenko@ucdconnect.ie (Illia Kovalenko), tpost@nu.edu.kz (Thierry Post)

1. Introduction

In security selection and asset allocation, a financial modeling approach based on Econometrics and Mathematical Programming often faces head winds in the form of a curse of dimensions, high sensitivity to estimation error and high transaction costs. Simple heuristic decision rules such as equal-weighted winners-minus-losers portfolios (Jegadeesh and Titman, 1993) and equal diversification (DeMiguel et al., 2009) are often superior out-of-sample. Improvement upon heuristics generally requires an eclectic approach which combines and integrates methods and techniques from Finance, Decision Science, Econometrics and Optimization.

This study develops a modeling approach to equity sector selection using Exchange Traded Funds (ETFs). An ETF is an investment fund that is traded on stock exchanges much like stocks. It generally operates with an arbitrage mechanism designed to keep it trading close to its net asset value, although deviations can occasionally occur. This particular application is promising for a number of reasons. The cross-section is limited to nine diverse sectors ETFs. Transaction costs are limited due to the low expense ratios and bid-ask spreads. Options are available for these ETFs, which allows for extracting forward-looking information about the return distribution from market prices and option valuation formulas.

Equity sector selection appears highly profitable using equal-weighted portfolios of stocks in winner industries (Moskowitz and Grinblatt, 1999). This approach is, however, based on paper portfolios which do not account for transactions costs. Earlier studies show limited evidence that this strategy can be applied in a profitable way due to high turnover (Andreu et al., 2013; Tse, 2015; Novy-Marx and Velikov, 2016; Vanstone et al., 2021). This study explores whether a financial modeling approach to sector ETF selection achieves significant risk-adjusted outperformance after transaction costs. To assess this, a new and comprehensive data set (2000-2020) is constructed of spots and options of sector ETFs that constitute the S&P500 market index. In addition, a number of techniques and methods are incorporated into a holistic methodology. These include:

- Forward-looking estimates for the marginal risk-neutral distribution (RND) of sector ETF returns from option prices;

- Parametric stochastic processes for extracting information from the options prices surface (all strikes and maturities);

- Risk transformation methods to obtain real-world distributions (RWDs) from RNDs;

- Multivariate copulas to model the mutual dependence between the sector ETFs;

- Monte Carlo simulation to draw samples and discretize the estimated joint RWD for numerical purposes;

- Winsorization of the simulated input samples to account for estimation and simulation error;

- Portfolio optimization with Stochastic Dominance (SD) constraints to account for the skewness and downside risk of the RWD of sector ETF portfolio returns.

The combined approach significantly outperforms the S&P500 index (SPX) and various simpler active approaches (including a sector-momentum heuristic and mean-variance optimization) in terms of a range of performance measures out of sample. All of the aforementioned techniques and methods are important for achieving outperformance, with the exception of the choice of the multivariate copula, which underlines the importance of using an integrated modeling approach.

Our work is related to several strands of literature. These include option-implied portfolio selection (Kostakis et al., 2011; DeMiguel et al., 2013; Kempf et al., 2015), copula-based portfolio optimization (Patton, 2004; Low et al., 2013; Kakouris and Rustem, 2014), portfolio optimization with SD constraints (Dentcheva and Ruszczyński, 2006; Roman et al., 2013; Post et al., 2018), SD-based option pricing (Constantinides et al., 2009; Constantinides et al., 2011; Post and Longarela, 2021), and robust portfolio management (see Fabozzi et al. (2007) for a review). However, to our knowledge, this study is the first to integrate all of the aforementioned methods and techniques into a feasible, coherent and effective methodology.

This study finds results of risk-adjusted outperformance, which add to the evidence against market efficiency. Given the low cost of ETFs, the modest rebalancing frequency and the publicly available data that we use, our strategy is applicable in a real-world environment and may be of interest to a wide range of investors and institutions.

The methodology is not limited to sector ETFs. Other possible applications that could benefit from the same methodology are the asset allocation across multiple currencies or commodity futures. In these cases, the dimensionality is limited, transactions costs are low and options are available.

2. Methodology

The portfolio optimization approach that is applied in this study can be bifurcated into two parts. First, we forecast the joint probability distribution of the 9 sector funds and the marginal probability distribution of the benchmark. Second, we conduct a numerical optimization of the fund portfolio using the predicted probabilities of the funds and benchmark returns. For each rebalancing date t, the estimation window I_t consists of the daily data of the previous 12 months in relation to spot and option prices. The forecast horizon and the holding period are both one month ($H = 1/12$ years).

Four steps are taken at every portfolio formation date (t): (1) estimate the marginal RWDs for every ETF and benchmark, (2) estimate the dependence structure of the funds' returns using copula models, (3) discretize the ETFs joint distribution using random sampling, and (4) optimize the portfolio with SD constraints. Table 1 provides the description of the competing investment strategies. A synopsis of these steps is given below; the technical details are presented in Supplementary Appendices A-E. Table 2 summarizes the notation that is used for variables and parameters.

[Insert Table 1]

[Insert Table 2]

2.1. Estimate the marginal RWDs of individual sector funds

Three alternative methods are used for estimating the marginal RWD of every sector ETF. The first two methods apply option-pricing models (BSM and Heston) to ETF option market prices. The third model (GJR-GARCH) uses a time-series estimation. While BSM assumes an i.i.d. process, Heston and GARCH are based on two different time-varying volatility processes. Given that volatility clustering is a well-known feature of financial data, the Heston and GARCH are expected to yield superior forecasts and portfolio performance.

The option-based approach yields RND estimates that need to be transformed into RWD estimates. One important shortcoming is that the RNDs have zero-valued risk premiums for all base assets, by definition. Additionally, the risk-neutral variance and kurtosis tend to be higher than the real-world counterparts. Furthermore, RNDs are more negatively skewed than real-world densities.

Bliss and Panigirtzoglou (2004) evaluate single-parameter transformations that take the shape of the Intertemporal Marginal Rate of Substitution of a single-parameter utility function of a representative agent. They find that ex-post, real-world densities for the S&P 500 and FTSE 100 indices are a significant improvement upon their risk-neutral densities. Liu et al. (2007) and Kostakis et al. (2011) provide further utility-based results for these markets. However, as empirical estimates of implied risk aversion are incompatible with a standard consumption-based framework (Jackwerth, 2000; Rosenberg and Engle, 2002; Ziegler, 2007), standard utility transformations are unlikely to provide completely satisfactory real-world densities.

Liu et al. (2007) and Shackleton et al. (2010) estimate the two parameters of a more flexible transformation, by maximizing the likelihood of the one-period-ahead equity prices. They demonstrate that this approach is superior to a one-parameter transformation in terms of fitting and forecasting the equity return distribution. We apply two established transformation methods here following Liu et al. (2007) and Shackleton et al. (2010). With regards to the lognormal distribution, we apply a statistical calibration using a two-parameter beta distribution as a calibration function.

For the Heston model, the application of a statistical beta calibration is complicated by the absence of an analytical cumulative distribution function (cdf) and inverse cdf. Therefore, we apply an alternative method by Atiya and Wall (2009) based on the adjustment of risk premia. Shackleton et al. (2010) show that this approach yields similar results as the two-parameter calibration when both the price and the variance equations are transformed. Therefore, this approach seems suitable for our Heston bivariate model. Simulation under the risk premium transformation does not require inverse cdf calculations and is based on a standard discretization scheme such as the Milstein one (Kahl and Jäckel, 2006).

In our application, the Atiya and Wall (2009) method adjusts the risk premiums directly using Maximum Likelihood estimates for the market price of risk specification $[\lambda_1\sqrt{(1-\rho^2)V_t}; \lambda_2\sqrt{V_t}]$ proposed by Aït-Sahalia and Kimmel (2007). While $\lambda_{1,2}$ are risk adjustment parameters which are given as a fraction of the variance V_t and ρ is the correlation between price and variance processes, the price risk premium equals $\lambda_1(1-\rho^2)+\lambda_2\rho$ reflecting its dependence on variance. This specification results in the bivariate real-world stochastic model given by Eq. (C2.1) in Appendix C2.

The steps for each model are as follows.

a. For the BSM model: 1) for each rebalancing date in the window I_t, find the end-of-day, nearest-the-money implied variance V for the nearest-to-expiry options; 2) using Eq. (B.3), calculate risk-neutral lognormal cdfs with forecast period H for all rebalancing dates in I_t (see Appendix B); 3) calculate the parameters a and b of the beta distribution in Eq. (C1.1) that is used as a calibration function by maximizing the likelihood of the one-period-ahead prices (see Appendix C1).

b. For the Heston model: 1) using the screened cross-section of options contracts on date t, calculate Heston model parameters, $\{V_0, \kappa, \theta, \sigma, \rho\}$, by minimizing mean squared error in Eq. (B.8) (see Appendix B); 2) using the estimation window I_t and calculated parameters, $\{V_0, \kappa, \theta, \sigma, \rho\}$, calculate risk-premium drifts, λ_1 and λ_2, for real-world dynamics in Eq. (C2.1) by maximizing likelihood in Eq. (C2.8) (see Appendices C2).

c. For the GJR-GARCH(1,1,1) model, using estimation window I_t, estimate parameter vector, $\{\omega,\alpha,\alpha^-,\beta,\upsilon\}$, by maximizing the likelihood of Eq. (E.2) (see Appendix E).

2.2. Estimate the between-fund dependence using copulas

Having estimated the marginal RWDs for every sector fund, we turn to the estimation of the joint distribution of the funds. Extracting the joint distribution from option prices requires options with multiple underlying assets that are not available for sector ETFs. Instead, we use the historical time series in the estimation window to estimate the dependence structure using copulas.

We use two multivariate copulas. The first one is a traditional Gaussian copula which can model only linear dependence using the Pearson correlation. The second is a Regular Vine (R-Vine) copula model, which allows modeling of nonlinear dependence between fund returns. Among other things, the R-Vine copula can model lower tail dependence typically seen in equity returns due to increased correlations in bad economic times (Ang and Chen, 2002; Patton, 2004; Low et al., 2013).

The R-Vine copula is based on the decomposition of the joint distribution into a product of conditional bivariate (pair) copulas. Each pair relationship can be modeled using various types of bivariate copulas.[1] The R-Vine copula is a graphical representation of the pair-copula construction in a sequence of 8 connected trees, where the number of trees is one less than the number of assets.

To fit the R-Vine copula, we use the Dissmann et al. (2013) algorithm (Algorithm 3.1 in Dissmann et al. (2013)). It comprises three components which are performed sequentially: (1) tree structure selection using a maximum spanning tree algorithm with weights being equal to Kendall correlation; (2) pair-copula selection from 49 bivariate copulas based on the Akaike criterion; (3) pair-copula parameters estimation using maximum likelihood estimator.

To obtain the copula estimates, we use the following two steps:

a. Extract the standardized residuals z of the estimated univariate GJR-GARCH models using Eq. (E.1) for the 9 sector ETFs using all observations in the estimation window I_t. Convert the standardized residuals z into pseudo-observations $u \in [0, 1]$ using the marginal probability distribution.

b. Fit the multivariate copula model (Gaussian or R-Vine) to the pseudo-observations, $\boldsymbol{u} = [u_1, \ldots, u_9]$, obtained in Step 2.2.a..

More details are provided in Appendix D.

[1] For our application, we use the following set of bivariate copulas: independence, Gaussian, Student-t, Gumbel, Clayton, Frank, Joe, BB1, BB6, BB7, BB8, Tawn (type 1 and 2) copulae and their rotated versions.

2.3. Draw a large input sample from the joint distribution

To discretize the joint distribution for numerical optimization, we use a Monte Carlo simulation to obtain a large input sample. The algorithms of the BSM, Heston and GARCH model sampling is described below in four steps:

a. Draw a sample of 10,000 pseudo-observations $\boldsymbol{u}$ from the estimated copula model, obtained in Step 2.2.b. (Algorithm 2.2. in Dissmann et al. (2013)).

b. For the BSM model: 1) using the calibration function parameters a and b estimated in Step 2.1.a., apply the inverse beta cdf to pseudo-observations $\boldsymbol{u}$ obtained in Step 2.3.a.; 2) using V found in Step 2.1.a., apply the inverse of lognormal cdf in Eq. (B.3) to the obtained values. This will yield a sample input from a joint probability distribution for the forecast period H.

c. For the Heston model: 1) using estimated parameters $\{V_0, \kappa, \theta, \sigma, \rho, \mu, \lambda_{1,2}\}$, obtained for 9 ETFs in Step 2.1.b., apply the discretization Milstein scheme to real-world price dynamic of each ETF in Eq. (C2.1) and simulate 10,000 returns for the forecast period H; 2) sort simulated returns according to copula sample $\boldsymbol{u}$, obtained in Step 2.3.a., to generate the joint distribution sample.

d. For the GJR-GARCH (1,1,1) model: 1) using the univariate GJR-GARCH models estimates, obtained for 9 ETFs in Step 2.1.c., simulate 10,000 returns for the forecast period H (see Appendix E); 2) sort simulated returns according to copula sample $\boldsymbol{u}$, obtained in Step 2.3.a., to generate the joint distribution sample.

2.4. Portfolio optimization with SD constraints

The optimal portfolio is required to dominate the benchmark by second-degree stochastic dominance (SSD) (Hadar and Russell, 1969; Hanoch and Levy, 1969; Rothschild and Stiglitz, 1970). SSD occurs if and only if all rational risk-averse investors with increasing and concave utility functions prefer the optimized portfolio to the benchmark. For non-Gaussian returns distributions, this stochastic order accounts for skewness and downside risk beyond variance. Using the sample input from section 2.3 with equal probabilities ($p_t = 10^{-5}$), we solve the SSD optimization problem that is specified in Appendix A. The optimal portfolio weights are then applied to holding period (H) returns. The framework can potentially be applied directly to higher degrees of SD (Post and Kopa, 2017; Fang and Post, 2022).

3. Data and Estimates

3.1. Data

The investment universe consists of nine sectors ETFs. The specific sector ETFs considered are Materials (XLB), Consumer Staples (XLP), Energy (XLE), Financial (XLF), Industrial (XLI), Technology (XLK), Utilities (XLU), Consumer Discretionary (XLY), and Health Care (XLV). These ETFs track the corresponding S&P 500 sector indices. Table 3 presents the assets under management, expense ratio and bid-ask spread of the ETFs. The median ETF bid-ask spread is 3 basis points, while the median expense ratio equals 0.19%.[2] The median market value of funds is $2,906 million.

The sample period is from January 2000 to December 2020. A one-year moving estimation window of daily returns is used for estimating the GARCH model parameters, risk transformations and copulas. The subperiod from January 2000 to December 2000 is used only as an estimation window and not as a portfolio holding period.

Previous studies indicate that option-implied measures work well in cases of short-term horizon forecasting. Specifically, Kostakis et al. (2011) and Kempf et al. (2015) use a monthly rebalancing frequency for the portfolio, while DeMiguel et al. (2013) consider three rebalancing intervals: daily, weekly, and fortnightly. We focus on a one-month holding period in this study and hence have 239 out-of-sample periods. The portfolio is rebalanced at the start of the first trading day of the month.

As a benchmark, we consider the S&P 500 equity index, (τ in Appendix A), the most widely used benchmark for US blue-chip stocks. We also consider the equally-weighted portfolio and an equal-weighted momentum portfolios comprising the one and three sectors with the highest mean return in the previous year.

[2]Further detail regarding SPDR S&P500 sector ETFs is available at https://www.spdrs.com/.

The options data were obtained from IvyDB US (OptionMetrics). The IvyDB US database contains a complete historical record of end-of-day data on all US exchange-traded equity and index options from January 1996 to December 2020. The S&P 500 index options are European style, while all the ETF options are American style. The historical data were obtained from the CRSP database. We analyze only the prices of OTM and ATM options because the ITM contracts are less actively traded and have higher early exercise premia.

Since a closed-form pricing formula is available only for European-type contracts and all ETF options are American style, we estimate equivalent European option prices from the American prices which have the same implied volatility, using the BSM model and the Barone-Adesi (1987) pricing formulae to estimate the early exercise premia. To minimize the early exercise premium, we use the OTM American options. Their effective maturity is closer to the maturity of the contract, and hence, only very small errors can be created by applying these formulae. Since the Heston model provides a closed-form formula for calls only, we obtain call option values from put options using the put-call parity.

Several screening conditions are used to eliminate illiquid options (Carr and Wu, 2009; Shackleton et al., 2010; Fan et al., 2018). First, we exclude option records when the bid price is zero, negative, or more than the ask price. Options with less than five days or more than one year until expiration are excluded from the sample. We also exclude options with less than five contracts traded over a day, and options contracts with zero open interest and/or volume. Options with implied volatilities bigger than 100% or lower than 1% are also excluded.

After the option screening, the average number of S&P 500 option prices per day is 818, made up of 588 puts and 230 calls. The average number of expiry dates available for the S&P 500 index options is 12. The ETF option datasets are significantly smaller. The average number of options is 74 contracts across all nine ETFs. The largest dataset available is for the XLE fund with 215 contracts per day (129 puts and 86 calls), while the smallest dataset is for the XLP fund with 40 contracts per day (27 puts and 13 calls). The average number of maturities available is 5.5 among all ETFs and does not vary significantly between funds. These screened data are used for the estimation of the option-implied marginal probabilities of the ETFs (base assets) and the S&P 500 index (benchmark) using the methodology described in Section 2.

Table 3 shows descriptive statistics for the daily historical returns data for the nine ETFs and the S&P 500 index for the whole sample period. The moments are calculated for each estimation window and medians and interquartile range of their distributions are provided. Most assets' distributions are negatively skewed, and assets exhibit excess kurtosis more than 75% of the time. Not surprisingly, a Gaussian distribution is rejected using the Jarque-Bera test at every conventional level of significance, for all assets. The non-normality of the data suggests that the classical mean-variance approach may lead to suboptimal portfolios. The stochastic dominance approach, in contrast, is a more general method, allowing asymmetric and heavy-tailed data.

In addition, we tested the data for ARCH and leverage effects. In Table 3, the ARCH coefficients are the autocorrelation coefficients for the squared returns computed for one lag. The positive values provide evidence of volatility clustering for all assets. The leverage coefficients are the linear correlations between the lagged returns, r_{t-1}, and the squared return, r^2, at time t. We observe the negative and statistically significant coefficients for all the ETFs and the index in more than 50% of periods. This results corresponds to the negative asymmetry in the marginal probability distributions.

[Insert Table 3 here]

Table 4 presents the unconditional Pearson correlation coefficients between each pair of ETFs. The table also shows the significant difference between the correlations of the S&P market index and its constituent ETFs when returns are below (above) the 25th (75th) percentile of the market daily returns. This pattern corroborates other studies (Ang and Chen, 2002; Patton, 2004; Low et al., 2013) that describe these conditional correlations as exceedance correlations. To account for this nonlinear dependence, we use the R-Vine copula. Unlike the Gaussian copula, vine copula is able to model asymmetric tail dependence.

[Insert Table 4 here]

3.2. Option-implied marginal densities

In this subsection, we investigate the option-implied RND, the effect of risk transformation and the predictive ability of the estimated RND and RWD. The emphasis will be on the

Heston model instead of the BSM, as the former is both more general and more effective for our purposes than the latter.[3]

Table 5 shows, for every sector fund, the median (across the formation months) of the central moments of three alternative one-month-ahead return distributions: empirical probability distribution (EPD), RND based on the Heston option pricing model and the RWD after the risk transformation. For comparison, results are also shown for the S&P500 index, the equally-weighted portfolio (EWP), the ETF with the highest mean return in the 12-month estimation window (Top-1 EPD), and the ETF with the highest mean according to the RWD (Top-1 RWD). The distribution estimate is significantly affected by the risk-premium transformation. As expected, the RWD mean and skewness are increased, while standard deviation and kurtosis declined. The last three columns of Table 5 reveal that risk transformation significantly improves the in-sample model fit, measured by the Log Likelihood Ratio (LLR) that is maximized to select the risk transformation parameters.

The Top-1 RWD ETF naturally shows the strongest divergence between RND and RWD moments. Since this portfolio has the most favorable mean and risk profile, it will play a prominent role in the portfolio optimization based on the RWD. By contrast, the RWD of the 'Top-1 EPD' ETF is similar to that of the median ETF and has a lower mean and skewness as well as higher volatility and kurtosis than Top-1 RWD. This result illustrates that sector fund selection based on the RWD generally does not resemble a standard momentum strategy.

[Insert Table 5 here]

The LLR ratio describes the improvement in the estimation window, and it may suffer from over-fitting to the data. To analyze the predictive ability of the RWD out of sample, we use the weighted likelihood ratio (WLR) test by Amisano and Giacomini (2007). Table 6 reports the values of the test statistic $\overline{WLR}$ which has an asymptotic standard normal distribution. The table shows that the RWD yields better forecasts than the RND for most sector funds, which mitigates concerns about over-fitting to the realized returns in the estimation window. Moreover, the RWD option-implied models (BSM and Heston) have higher predictive ability than the GARCH model. Since the RWD moves probability mass from the

[3] Additional details of the results for the BSM are available upon request from the authors.

left to the right of the support of the distribution, it is not surprising that it achieves a lower likelihood ratio for the left tail, for some of the funds.

[Insert Table 6 here]

3.3. Joint density estimation using copulas

We now investigate the multivariate copula fit. Table 7 reports the median (across the formation months) of the values of several dependence and model fit measures implied by the Gaussian and R-Vine copula (see Appendix D). For each ETF, we averaged the values of the pairwise measures across all 8 pairs of ETFs. For comparison, results are shown also for the S&P500 index, the equally-weighted portfolio (EWP), the ETF with the highest mean return in the 12-month estimation window (Top-1 EPD), and the ETF with the highest mean according to the RWD (Top-1 RWD).

[Insert Table 7 here]

As dependence measures, we use the unconditional linear correlation and the difference between the Ang and Chen (2002) exceedence correlation calculated at 25th and 75th percentiles. These dependence measures are calculated using a sample of 10,000 pseudo-observations u drawn from the estimated copula (Step 2.3.a.). The values are lower than those provided in Table 4 because $\boldsymbol{u}$ is not affected by marginals and outliers, unlike empirical returns data. We can observe that the two copula models imply very similar Pearson correlation levels for all funds. This pattern suggests that the choice of the kernel model will not materially affect the Mean-Variance optimization of sector fund portfolios.

Since the Gaussian copula is symmetric, it implies zero exceedance correlation difference. In the case of the R-Vine copula, we can observe a moderate dependence asymmetry which is in line with previous studies which document increased dependence during down markets (Ang and Chen, 2002; Patton, 2004; Low et al., 2013). The elevated left tail correlations can be expected to discourage portfolio diversification based on SSD optimization, if the joint distribution is estimated using the R-Vine copula.

As model fit measures, we use the log-likelihood, the Akaike and Schwarz (Bayesian) criterion. We can observe that the R-Vine copula provides a better fit to the data than the Gaussian does for all sector funds. Even after adjusting for the number of parameters and estimation error, the R-Vine model remains superior, as can be witnessed by the better (lower) scores on the Akaike and Schwarz criteria. Whether the superior model fit affects the portfolio composition and investment performance of optimized sector fund portfolios of course remains to be seen; this will depend, among other things, on whether or not the dominance constraints are binding in the tails of the distribution.

4. Portfolio performance analysis

In this section, we analyze the out-of-sample performance of the ten competing portfolios described in Table 1. Two strategies are passive: the benchmark S&P500 index and the equally-weighted portfolio. Eight portfolios are constructed from the nine sector ETFs to enhance the S&P500 benchmark index. The portfolio rebalancing takes place monthly on the first trading day of the month from January 2001 to December 2020. Performance evaluation is based on monthly returns during the subsequent holding period.

4.1. Evaluation methods

The performance evaluation uses the first four central moments to describe the location, dispersion and shape of the portfolio return distribution. As economic risk measures, the value at risk (VaR) and conditional value at risk (CVaR) for 95% confidence level are included. As performance measures, we consider the certainty equivalent (CE) return for a power utility function with RRA = 4, $u(x) = \frac{x^{-3}-1}{-3}$, the Sharpe ratio and the Sortino ratio. A number of statistical tests are performed to determine statistical significance.

A paired t-test is used to determine the statistical significance of the mean spread between a given portfolio and the benchmark. We use the statistical bootstrap test by Ledoit and Wolf (2008) to test the null hypothesis that the active portfolio has the same Sharpe ratio as the benchmark. The statistical resampling test by Linton et al. (2005) is used to test the hypothesis that the active portfolio stochastically dominates the benchmark out of sample. Turnover serves as a proxy for the level of transaction costs associated with portfolio rebalancing. It is computed as the average absolute change summed across all 9 ETF weights over 239 out-of-sample periods:

$$\text{Turnover } = \frac{1}{239}\sum_{t=1}^{239}\sum_{i=1}^{9}\left|\omega_{i,t+1}-\omega_{i,t+}\right|$$

where $\omega_{i,t+1}$ is the optimal portfolio weight of asset i for period t+1 and $\omega_{i,t+}$ is the portfolio weight of asset i before rebalancing at time t+1. Additionally, we calculate the portfolio performance net of transaction costs. To assess it, let pc be the proportional transaction cost. We allow the transaction costs to vary in time and between base assets, $pc_{i,t}$. When a portfolio is rebalanced, the total proportional cost is given by $\sum_{i=1}^{N} pc_{i,t+1}\left|\omega_{i,t+1}-\omega_{i,t+}\right|$. The evolution of wealth net of transaction costs (NW) for portfolio ω is given by

$$NW_{t+1} = NW_t\left(1+\boldsymbol{R_t^T}\boldsymbol{\omega}\right)\left[1-\sum_{i=1}^{N} pc_{i,t+1}\left|\omega_{i,t+1}-\omega_{i,t+}\right|\right]$$

The terminal wealth (NW_T) net of transaction costs will be equal to:

$$NW_T = NW_1\left(\prod_{t=1}^{1}\left(1+\boldsymbol{R}_t^T\boldsymbol{\omega}\right)\left[1-\sum_{i=1}^{N} pc_{i,t+1}\left|\omega_{i,t+1}-\omega_{i,t+}\right|\right]\right).$$

Specifically, we model terminal wealth by hypothetically investing \$100 (NW_1) at the start of the out-of-sample periods for each portfolio management strategy. The return net of transaction costs, $RNTC_{t+1}$, for any strategy at time $t+1$ is then given by

$$RNTC_{t+1} = \frac{NW_{t+1}}{NW_t} - 1$$

We then calculate other performance metrics using the return net of transaction costs, $RNTC$. We use ETFs bid-ask spreads recorded on the rebalancing dates as proportional transaction costs.

4.2. Base case

Table 8 summarizes the out-of-sample performances of the competing portfolio strategies. The benchmark index (S&P 500 in Column 1) on average yields 0.627% per month, with a standard deviation of 5.32%, while its negative skewness of -1.43 reflects the effects of stochastic market volatility and stochastic between-stock correlation.

The equally-weighted passive strategy (EWP in Column 2) outperforms the benchmark index across all considered performance measures, enhancing its mean return by 245 basis points (bps) per annum. The improvements are both economically and statistically significant according to all considered tests. This result is expected given earlier studies demonstrating the robust performance improvement of EWP (DeMiguel et al., 2009).

In Columns 3–5 of Table 8, we provide the performance of portfolios selected based on historical data. The ETF with the highest mean return in the 12-month estimation window (Top-1), which tries to capture the industry momentum effect, enhances the mean return of the SPX by 297 bps per annum. This improvement is only slightly higher than what is achieved using the simple EWP. In addition, the Sharpe ratio is not significantly improved, based on the Ledoit and Wolf (2008) test. The equally-weighted winners portfolio (Top-3), demonstrates similar results to Top-1 but with lower volatility due to better diversification. The active strategy based on the GJR-GARCH model performs on par with EWP and momentum strategies but has a lower CVaR95% value. The p-values for Ledoit and Wolf (2008) and mean spread tests, however, reveal the statistical insignificance of GARCH improvement relative to the SPX.

In Columns 6–11 of Table 8, the option-implied strategies reveal substantial additional performance improvements. When we estimate the inputs using the BSM marginal model, the results trail the heuristic EWP and momentum strategies. This result is consistent with the inability of the model to account for the negative skewness and excess kurtosis of the risk profile of the S&P500 index and sector ETFs (see Table 3).

[Insert Table 8 here]

By contrast, remarkable performance improvements are found when the optimization uses option-implied probabilities, based on the Heston model, which accounts for time-varying variance. Using the Gaussian copula, the Heston portfolio achieves a mean return of 1.23% per month, which is an annualized 7.20% increase compared to the S&P 500. Despite the higher standard deviation of 5.13%, the risk-adjusted performance is superior to the benchmark and the heuristic EWP and Top-3 strategies, with a CE of 0.46, Sharpe ratio of 0.20 and Sortino ratio of 0.29. The Sharpe ratio increase of 104% is statistically significant at the 5% level; the hypothesis of out-of-sample stochastic dominance cannot be rejected (p-value: 0.005).

The copula structure does not seem essential here. Using the R-Vine copula to model the dependence structure does not yield any improvement compared to the Gaussian copula, both for the BSM (not tabulated) and the Heston model (Column 7 in Table 8), despite the compelling evidence for non-linear dependence between the sector funds. A comparison of the distribution of the SPX and the optimal portfolio reveals that the invariance to the copula structure and tail risk estimation is related to the favorable mean and skewness of the optimal portfolio. Due to the improvements of the mean and skewness, the benchmark risk restrictions are not binding in the top and bottom deciles, for more than 75 percent of sample.

Using winsorized sample input yields more substantial improvement, especially in terms of downside risk. Wins has the least negatively skewed return distribution and it maximizes CE, Sharpe and Sortino across all specifications. The winsorization, however, is not essential to outperforming the passive and heuristic strategies, and the incremental effect is smaller than that of the adoption of the Heston model.

The use of the SD order is essential to achieve these attractive results. If we use the mean-variance approach, the performance of the Heston-based portfolio materially deteriorates: CE is 0.29, Sharpe is 0.15 and Sortino is 0.21. While the portfolio still outperforms the stock index and heuristic strategies, the numbers for MVD are well below those for SD. The superior performance of SD compared with MVD is consistent with Post and Kopa (2017), Post et al. (2018), and Fang and Post (2022). It reflects that SD can achieve a better balance between upside potential and downside risk by considering the entire risk profile of portfolios instead of the variance only.

The risk transformation is even more crucial for achieving outperformance. Heston/RND significantly underperforms the strategies based on real-world Heston probabilities (Heston, RVine, Wins and MVD), because all ETFs have the same expected return under the RND. The risk transformation translates the differences between the ETFs in the dispersion and shape of the RND to differences in RWD expected return.

Figure 1 shows the cumulative returns of the competing strategies. The relative performance observed in Table 8 is reflected in the graph. Heston and Wins lead all other portfolios from 2001 until 2020.

[Insert Figure 1 here]

The Heston-based strategies yield the highest average turnover, which means that they are more exposed to transaction costs than the other portfolios. This, in particular, can be due to the high sector concentration of Heston-based strategies, which have the highest values of the squared sum of weights among all active strategies. To account for this effect directly, we also calculate the portfolios' performance net of proportional transaction costs ($pc_{i,t}$). The results are presented in Table 9. As expected, we observe the most significant reduction of profits for the Heston-based portfolios. Heston and Wins, however, still provide significant outperformance of the benchmark index and improvements upon the heuristics strategies and the alternative active strategies. EWP, which is the least affected by transaction costs, also beats S&P 500 but underperforms Heston and Wins in terms of risk-adjusted performance.

The after-costs results remain strong: for the Heston portfolio, a CE of 0.29, Sharpe ratio of 0.17 and Sortino ratio of 0.24. Again, these improvements stem from the combined use of the Heston model and SD order; further improvements can be achieved using winsorization.

[Insert Table 9 here]

Here, we conclude that portfolio optimization with option-implied data offers superior forecasting probabilities and can lead to better out-of-sample performance of the corresponding portfolios. Regarding the marginal distribution, the Heston model yields the best performance in terms of risk-adjusted returns, while the BSM model cannot outperform passive strategies and the portfolios constructed using the historical distribution. The Heston model, however, has the highest turnover requirements. This can be explained by the trade-off between portfolio performance and a necessity to rebalance the portfolio more regularly, revealed in Carroll et al. (2017). Despite the high turnover, the Heston model yields better portfolio performance than the benchmark, EWP and other approaches, even after accounting for transaction costs. Winsorizing improves the asymmetric risk profile of the Heston-based portfolio, and mean-variance is inferior to SSD when applied to the sector ETFs. The risk-neutral probability yields poor out-of-sample performance showing the importance of the risk transformation for proper portfolio selection. Furthermore, the results show that accounting for non-linear dependence (using the R-Vine copula) does not yield any gains which are statistically or economically significant.

4.3. Portfolio performance during change of market conditions

Option-implied probability estimates rely on the current cross-section of option prices. Option-implied strategies should yield better performance during crisis periods for two reasons. First, the flow of information is higher during bad economic times, such as during crises, or during highly volatile periods. In such market conditions, historical information becomes rapidly out-of-date. Therefore, it is less valuable than the option-implied estimator, which relies solely on current data and more accurately reflects prevailing market conditions. Second, option markets offer more benefits for informed investors than spot stock markets (Easley et al., 1998; Pan and Poteshman, 2006; Hu, 2011). Therefore, one expects to observe more informed investors in the option market during periods of higher information asymmetry. This would make option prices more informative, again suggesting that the option-based portfolios should yield significantly superior performance in crisis periods. In non-crisis periods, however, when fewer extreme events occur, and the information asymmetry is lower, option-implied probabilities may have less benefit over historical ones. Meanwhile, several studies find that option-implied distributions do not anticipate stock market crashes (Bates, 1991; Gemmill and Saflekos, 2000). We test this hypothesis in this subsection.

To test this hypothesis, we compare performances during periods of high and low market volatility. As a proxy for volatility, we use the VIX (CBOE Volatility) index, which represents the market's expectation of monthly forward-looking volatility. The VIX index is widely used as an indicator of market risk and investor sentiments in industry and the financial literature. High VIX values are generally linked to large volatility resulting from increased uncertainty, risk, and investors' fear, while low VIX values correspond to stable, stress-free periods in the markets. The low VIX period is identified as the months when the VIX is below the 25th percentile of the VIX values distribution. Correspondingly, the high VIX period is identified as the months when the VIX is above the 75th percentile of the VIX values distribution.

Table 10 shows the results for high VIX (Panel A) and low VIX (Panel B). In the period of high volatility, most strategies yield a negative average return except Heston, RVine, Wins and MVD. Therefore, it is not meaningful to compare portfolios by the Sharpe and Sortino ratios (see Israelsen (2005) for an explanation of the possible paradoxes). Therefore, we calculated p-values for tests of significance differences between the benchmark and SSD portfolios, separately, for the mean return and standard deviation. We reject the null of a mean return difference for these Heston-based portfolios, while Wins also maximizes skewness and CE returns. When the market is less volatile, none of the portfolios can generate economically or statistically significant gains for the investor relative to S&P 500. When the market has moderate volatility (not tabulated), the Heston-based portfolios yield superior performance. These results are not surprising since the Heston-based portfolios are formed based on the stochastic volatility model. Therefore, the investor will benefit from them most during periods of high VIX, when an accurate volatility forecast is more crucial than during stress-free periods in the market. It is also consistent with recent findings in the literature that the stock volatility predictability is state-dependent, and strongest during bad economic times when the market is highly volatile (see Li and Zakamulin (2020) and references therein).

[Insert Table 10 here]

The market volatility is strongly correlated with momentum factor crashes (Barroso and Santa-Clara, 2015; Daniel and Moskowitz, 2016). Therefore, the Heston model allows to predict this factor crash and avoid its negative consequences for portfolio performance. Specifically, if the estimation window includes factor crashes, the defensive sectors which are less correlated with the momentum factor will appear attractive. Given a strong mean-reversion effect, including these ETFs in a portfolio would then lower the portfolio's expected return (Post et al., 2018). In the investigated sample, the most severe momentum crash corresponds to the estimation period from December 2008 to November 2009, with a mean monthly return of -5.76%. The Heston-based portfolios maximize the following month return with a value of 8.44%. GARCH provides a performance less than half of the Heston approach, with 3.71% return, while EWP and Top-3 yield 1.75% and 2.08% returns, respectively. BSM's performance is inferior with less than 1.00% return in December 2009.

5. Conclusion

Sector ETFs seem an appealing instrument for equity sector allocation and rotation. ETFs allow for diversification and rebalancing at a low cost. Moreover, the liquid sector-

ETF options market provides forward-looking information about the returns distribution.

We adopt a comprehensive financial modeling approach to sector investing with ETFs. Significant and robust outperformance is achieved. The financial modeling approach beats the passive benchmark and simpler active approaches (including a pure sector momentum play) in terms of a range of performance measures out-of-sample. The outperformance of the optimal portfolios is more pronounced in high volatility states, when markets are expected to be least efficient and option prices are expected to be most informative.

Several elements of the methodology stand out as being particularly effective: option-implied probabilities estimated using the Heston stochastic volatility model; risk transformation of the risk-neutral distribution; winsorizing of the Monte Carlo input sample; and the use of SD to account for skewness and tail risk.

The use of non-Gaussian multivariate copulas adds limited benefit, despite the strong evidence for non-linear dependence between the sector ETFs. The invariance to the copula structure stems from the benchmark risk restrictions being non-binding in the tails of the distribution. This is caused by the fact that the optimal portfolio has more favorable mean and skewness than the market index.

The performance of the RWD estimate that is based on the Heston model and risk-transformation methods is supported by three pieces of evidence: (i) the LLR test statistic in the estimation window for the risk premiums; (ii) the WLR test statistic for out-of-sample predictive ability; (iii) the outperformance of portfolios formed based on the RWD.

The results show that a financial modeling approach can outperform heuristic approaches, at least in our benign environment with a narrow cross-section of diversified base assets (9 sector ETFs) and the availability of exchange-traded options and low transactions costs, provided that insights and methods from Finance, Decision Science, Econometrics and Optimization are properly integrated.

The results can also be seen as adding to the evidence against market efficiency. Given the low costs of ETFs, a modest rebalancing frequency and the publicly available data that we use, our strategy is applicable in a real-world environment and for a wide range of investors and institutions.

We conclude with two promising areas for further research. First, despite assuming global risk aversion, the SSD does not assume skewness preference and kurtosis aversion, which has the effect of limiting the admissible set of sector fund portfolios. To exclude unrealistic risk preferences, the optimization over the forward-looking RWD could be extended to higher orders of SD and approximate SD orders, as in Fang and Post (2022). This approach is however computationally demanding, because our method uses very large pseudo-samples from the RWD, and the additional computational burden of using higher-order SD increases quickly with the sample size. Further research could focus on reducing the computer processing burden using problem reduction, tight sufficient conditions for dominance and/or tailor-made non-smooth optimization algorithms.

Second, the methodology for RWD estimation and SSD optimization can be extended to other universes of base assets, for example, currency pairs and commodity futures. Also in these cases, transactions costs are low and exchange-traded options are available. The choice of the copula model may be more relevant in those cases than for equity sector funds, as the diversification of tail risk and the estimation of tail dependence are expected to play a bigger role. If the number of base assets is much larger than in our application to 9 sector funds, the curse of dimension may complicate the dependence structure modeling. We can address this issue by using copula models that account for high multidimensionality, in particular, the truncated vine copula Brechmann et al. (2012) and factor copula (Oh and Patton, 2017).

Acknowledgements

The authors acknowledge the support of Science Foundation Ireland under Grant Number 16/SPP/3347.

References

Aït-Sahalia, Y. and Kimmel, R. (2007). Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics*, 83(2):413–452. 2.1

Amisano, G. and Giacomini, R. (2007). Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business & Economic Statistics*, 25(2):177–190. 3.2, 6

Andreu, L., Swinkels, L., and Tjong-A-Tjoe, L. (2013). Can exchange traded funds be used to exploit industry and country momentum? *Financial Markets and Portfolio Management*, 27(2):127–148. 1

Ang, A. and Chen, J. (2002). Asymmetric correlations of equity portfolios. *Journal of Financial Economics*, 63(3):443–494. 2.2, 3.1, 3.3, 7

Atiya, A. and Wall, S. (2009). An analytic approximation of the likelihood function for the heston model volatility estimation. *Quantitative Finance*, 9(3):289–296. 2.1, 1

Barone-Adesi, G. . W. R. E. (1987). Efficient analytic approximation of american option values. *The Journal of Finance*, 42(2):301–320. 3.1

Barroso, P. and Santa-Clara, P. (2015). Momentum has its moments. *Journal of Financial Economics*, 116(1):111–120. 4.3

Bates, D. (1991). The crash of 87: was it expected? the evidence from options markets. *The Journal of Finance*, 46(3):1009–1044. 4.3

Bliss, R. and Panigirtzoglou, N. (2004). Option-implied risk aversion estimates. *The Journal of Finance*, 59(1):407–446. 2.1

Brechmann, E., Czado, C., and Aas, K. (2012). Truncated regular vines in high dimensions with application to financial data. *The Canadian Journal of Statistics*, 40(1):68–85. 5

Carr, P. and Wu, L. (2009). Variance risk premiums. *Review of Financial Studies*, 22(3):1311–1341. 3.1

Carroll, R., Conlon, T., Cotter, J., and Salvador, E. (2017). Asset allocation with correlation: A composite trade-off. *European Journal of Operational Research*, 262(3):1164–1180. 4.2

Constantinides, G., Czerwonko, M., Jackwerth, J., and Perrakis, S. (2011). Are options on index futures profitable for risk-averse investors? empirical evidence. *The Journal of Finance*, 66(4):1407–1437. 1

Constantinides, G., Jackwerth, J., and Perrakis, S. (2009). Mispricing of s&p 500 index options. *The Review of Financial Studies*, 22(3):1247–1277. 1

Daniel, K. and Moskowitz, T. (2016). Momentum crashes. *Journal of Financial Economics*, 122(2):221–247. 4.3

DeMiguel, V., Garlappi, L., and Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies*, 22(5):1915–1953. 1, 4.2

DeMiguel, V., Plyakha, Y., Uppal, R., and Vilkov, G. (2013). Improving portfolio selection using option-implied volatility and skewness. *Journal of Financial and Quantitative Analysis*, 48(6):1813–1845. 1, 3.1

Dentcheva, D. and Ruszczyński, A. (2006). Portfolio optimization with stochastic dominance constraints. *Journal of Banking & Finance*, 30(2):433–451. 1

Dissmann, J., Brechmann, E., Czado, C., and Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, 59:52–69. 2.2, a.

Easley, D., O'hara, M., and Srinivas, P. (1998). Option volume and stock prices: Evidence on where informed traders trade. *The Journal of Finance*, 53(2):431–465. 4.3

Fabozzi, F., Kolm, P., Pachamanova, D., and Focardi, S. (2007). *Robust portfolio optimization and management*. John Wiley & Sons, Hoboken, New Jersey, 1st edition. 1

Fan, R., Taylor, S., and Sandri, M. (2018). Density forecast comparisons for stock prices, obtained from high-frequency returns and daily option prices. *Journal of Futures Markets*, 38(1):83–103. 3.1

Fang, Y. and Post, T. (2022). Optimal portfolio choice for higher-order risk averters. *Journal of Banking & Finance*, page 106429. 2.4, 4.2, 5

Gemmill, G. and Saflekos, A. (2000). How useful are implied distributions?: Evidence from stock index options. *The Journal of Derivatives*, 7(3):83–91. 4.3

Glosten, L., Jagannathan, R., and Runkle, D. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801. 2

Hadar, J. and Russell, W. (1969). Rules for ordering uncertain prospects. *The American Economic Review*, 59(1):25–34. 2.4

Hanoch, G. and Levy, H. (1969). The efficiency analysis of choices involving risk. *Review of Economic Studies*, 36:335–346. 2.4

Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2):327–343. 2

Hu, J. (2011). Does option trading convey stock price information? *Journal of Financial Economics*, 111(3):625–645. 4.3

Israelsen, C. (2005). A refinement to the sharpe ratio and information ratio. *Journal of Asset Management*, 5(6):423–427. 4.3

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance*, 48(1):65–91. 1

Kahl, C. and Jäckel, P. (2006). Fast strong approximation monte carlo schemes for stochastic volatility models. *Quantitative Finance*, 6(6):513–536. 2.1

Kakouris, I. and Rustem, B. (2014). Robust portfolio optimization with copulas. *European Journal of Operational Research*, 235(1). 1

Kempf, A., Korn, O., and Saßning, S. (2015). Portfolio optimization using forward-looking information. *Review of Finance*, 19(1):467–490. 1, 3.1

Kostakis, A., Panigirtzoglou, N., and Skiadopoulos, G. (2011). Market timing with option-implied distributions: A forward-looking approach. *Management Science*, 57(7):1231–1249. 1, 2.1, 3.1

Ledoit, O. and Wolf, M. (2008). Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance*, 15(5):850–859. 4.1, 4.2, 8, 9

Li, X. and Zakamulin, V. (2020). Stock volatility predictability in bull and bear markets. *Quantitative Finance*, 20(7):1149–1167. 4.3

Linton, O., Maasoumi, E., and Whang, Y.-J. (2005). Consistent testing for stochastic dominance under general sampling schemes. *The Review of Economic Studies*, 72(3):735–765. 4.1, 8, 9

Liu, X., Shackleton, M., Taylor, S., and Xu, X. (2007). Closed-form transformations from risk-neutral to real-world distributions. *Journal of Banking & Finance*, 31(5):1501–1520. 2.1

Low, R., Alcock, J., Faff, R., and Brailsford, T. (2013). Canonical vine copulas in the context of modern portfolio management: Are they worth it? *Journal of Banking & Finance*, 37(8):3085–3099. 1, 2.2, 3.1, 3.3

Moskowitz, T. and Grinblatt, M. (1999). Do industries explain momentum? *The Journal of Finance*, 54(4):1249–1290. 1

Novy-Marx, R. and Velikov, M. (2016). A taxonomy of anomalies and their trading costs. *The Review of Financial Studies*, 29(1):104–147. 1

Oh, D. and Patton, A. (2017). Modeling dependence in high dimensions with factor copulas. *Journal of Business & Economic Statistics*, 35(1):139–154. 5

Pan, J. and Poteshman, A. M. (2006). The information in option volume for future stock prices. *The Review of Financial Studies*, 19(3):871–908. 4.3

Patton, A. (2004). On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *Journal of Financial Econometrics*, 2(1):130–168. 1, 2.2, 3.1, 3.3

Post, T., Karabati, S., and Arvanitis, S. (2018). Porstfolio optimization based on stochastic dominance and empirical likelihood. *Journal of Econometrics*, 206(1):167–186. 1, 4.2, 4.3

Post, T. and Kopa, M. (2017). Portfolio choice based on third-degree stochastic dominance. *Management Science*, 63(10):3381–3392. 2.4, 4.2

Post, T. and Longarela, I. (2021). Risk arbitrage opportunities for stock index options. *Operations Research*, 69(1):100–113. 1

Roman, D., Mitra, G., and Zverovich, V. (2013). Enhanced indexation based on second-order stochastic dominance. *European Journal of Operational Research*, 228(1):273–281. 1

Rothschild, M. and Stiglitz, J. (1970). Increasing risk: I. a definition. *Journal of Economic Theory*, 2(3):225–243. 2.4

Shackleton, M., Taylor, S., and Yu, P. (2010). A multi-horizon comparison of density forecasts for the S&P 500 using index returns and option prices. *Journal of Banking & Finance*, 34(11):2678–2693. 2.1, 3.1

Tse, Y. (2015). Momentum strategies with stock index exchange-traded funds. *The North American Journal of Economics and Finance*, 33:134–148. 1

Vanstone, B., Hahn, T., and Earea, D. (2021). Industry momentum: an exchange-traded funds approach. *Accounting & Finance*, 61(3):4007–4024. 1

Table 1: Competing investment strategies

Name	Data	Marginal probability	Dependence model	Objective	Stochastic order	Winsorization	Risk transformation
S&P500 (SPX)	-	-	-	-	-	-	-
EWP	-	-	-	-	-	-	-
Top 1 (Top-3)	Historical	-	-	Max Mean	-	-	-
GARCH	Historical	GJR-GARCH	Gaussian	Max Mean	SSD	-	-
BSM	Options	BSM model	Gaussian	Max Mean	SSD	-	Statistical calibration
Heston	Options	Heston model	Gaussian	Max Mean	SSD	-	Atiya and Wall (2009) method
RVine	Options	Heston model	R-Vine	Max Mean	SSD	-	Atiya and Wall (2009) method
Wins	Options	Heston model	R-Vine	Max Mean	SSD	Yes	Atiya and Wall (2009) method
MVD	Options	Heston model	Gaussian	Max Mean	MVD	-	Atiya and Wall (2009) method
Heston RND	Options	Heston model	Gaussian	Min Variance	-	-	No

The table describes ten competing strategies. First column: short names of the portfolios. Second column: type of data used for the estimation of the marginal probabilities: option prices or historical returns. Third column: parametric model for the marginal distributions: Black-Scholes-Merton model ('BSM'), Heston stochastic volatility model indicates ('Heston') or GJR-GARCH (1,1,1) model with Student's t innovations ('GJR-GARCH'). Fourth column: model used to measure the dependence between the base assets: 'Gaussian' or 'R-Vine'. Fifth column: objective function used for portfolio optimization. Sixth column: stochastic ordering used for portfolio selection: second-degree stochastic dominance ('SSD') or Mean-Variance dominance ('MVD'). The final columns indicate whether the Monte-Carlo input sample of ETFs and S&P 500 index return is winsorized at 5th and 95th percentiles, and whether the risk transformation is applied to the marginal distribution.

Table 2: Summary of Notation

Notation	Description
ETF	exchange-traded fund
RND	risk-neutral probability distribution
RWD	real-world probability distribution
t	rebalancing date
I_t	estimation window
H	investment horizon
BSM model	Black-Scholes-Merton model of price S: $dS/S = (r-q)dt + \sqrt{V}dW$
r	risk-free rate
q	dividend yield
V	variance
a, b	shape parameters of beta calibration density: $u^{a-1}(1-u)^{b-1}/B(a,b), 0 \leq u \leq 1$
u	cumulative distribution function (cdf)
$B(a,b)$	beta function
Heston model	Heston (1993) stochastic volatility model:
	$dS/S = (r - q + (\lambda_1(1-\rho^2) + \lambda_2\rho)V)\, dt + \sqrt{V}d\, W_1$
	$dV = (\kappa\theta - (\kappa - \lambda_2\sigma)V)dt + \sigma\sqrt{V}d\, W_2$
	$\text{cor}(\, W_1,\ W_2) = \rho$
V_0	initial variance value
θ	long-term variance towards which the stochastic variance, V, reverts
ρ	correlation between the two Wiener processes, W_1 and W_2
κ	rate of reversion of V toward θ
σ	volatility of the volatility parameter
$\lambda_{1,2}$	risk premium parameters
GJR-GARCH model	asymmetric conditional heteroskedasticity (Glosten et al., 1993) model:
	$V_{t+1} = \omega + \alpha\varepsilon_t^2 + \alpha^- \max(0, -\varepsilon_t)^2 + \beta V_t$
ω	constant variance term
α	coefficient of the ARCH term
α^-	leverage (asymmetry) term coefficient
β	coefficient of the GARCH term
ε_t	error term, $\varepsilon_t = \sqrt{V_t}z_t$
z_t	standardized residuals, $z \sim \textit{Student-t}(\nu)$
ν	degrees of freedom for the Student-t marginal
$\boldsymbol{u}$	multivariate cdf sample (pseudo-observations), $\boldsymbol{u} = [u_1, \ldots, u_9]$, where $u \in [0,1]^N$
R-Vine copula	regular vine copula comprising pair-copulas of 49 types
p_t	equal probabilities applied to sample of 10,000 observations for portfolio optimization

Table 3: Descriptive statistics for daily return series

	Mean	Std. Dev.	Skewness	Kurtosis	ARCH	Leverage	Jarque-Bera	AUM	Expense ratio	BA Spread
					Median					
S&P 500	0.03	0.97	-0.16	4.37	0.08	-0.10	0.00			
Materials (XLB)	0.04	1.18	-0.22	3.95	0.09	-0.08	0.01	1797.27	0.19	0.03
Consumer Staples (XLP)	0.04	0.77	-0.20	4.25	0.11	-0.07	0.00	2906.40	0.19	0.03
Energy (XLE)	0.03	1.55	-0.28	3.99	0.09	-0.09	0.00	6203.74	0.19	0.02
Financial (XLF)	0.04	1.18	-0.14	4.19	0.11	-0.10	0.00	6566.80	0.19	0.05
Industrial (XLI)	0.04	1.08	-0.16	4.16	0.07	-0.07	0.00	2722.12	0.19	0.03
Technology (XLK)	0.04	1.17	-0.16	4.44	0.07	-0.08	0.00	4550.25	0.19	0.04
Utilities (XLU)	0.04	0.96	-0.26	4.19	0.14	-0.07	0.00	3607.95	0.19	0.03
Consumer Discretionary (XLY)	0.05	1.10	-0.21	4.29	0.09	-0.08	0.00	1815.59	0.19	0.03
HealthCare (XLV)	0.03	0.94	-0.21	4.34	0.12	-0.07	0.00	2709.91	0.19	0.03
Median ETF	0.04	1.10	-0.21	4.19	0.09	-0.08	0.00	2906.40	0.19	0.03
EWP	0.04	0.92	-0.25	4.45	0.09	-0.11	0.00	3578.64	0.19	0.03
					IQR					
S&P 500	0.07	0.63	0.47	1.98	0.24	0.11	0.01			
Materials (XLB)	0.08	0.74	0.40	0.93	0.22	0.12	0.03	2366.63	0.10	0.08
Consumer Staples (XLP)	0.04	0.34	0.41	1.29	0.21	0.15	0.02	7186.73	0.10	0.10
Energy (XLE)	0.14	0.60	0.39	1.63	0.14	0.12	0.06	8583.78	0.10	0.05
Financial (XLF)	0.10	0.96	0.37	1.76	0.23	0.12	0.03	16493.37	0.10	0.04
Industrial (XLI)	0.08	0.74	0.47	1.35	0.17	0.10	0.03	7402.22	0.10	0.10
Technology (XLK)	0.09	0.70	0.57	1.60	0.17	0.10	0.02	11705.17	0.10	0.10
Utilities (XLU)	0.06	0.42	0.51	2.35	0.17	0.11	0.05	4967.21	0.10	0.08
Consumer Discretionary (XLY)	0.06	0.78	0.53	1.72	0.16	0.09	0.04	9747.37	0.10	0.10
HealthCare (XLV)	0.06	0.45	0.42	2.25	0.16	0.14	0.02	11485.47	0.10	0.08
Median ETF	0.08	0.70	0.42	1.63	0.17	0.12	0.03	8583.78	0.10	0.08
EWP	0.06	0.52	0.41	1.96	0.23	0.11	0.01	8852.17	0.10	0.08

This table presents descriptive statistics for the daily returns of nine sectors ETFs and the S&P 500 index. The statistics are calculated for each estimation window and medians and interquartile ranges (IQR) of the distributions are provided. The full sample runs from January 2000 to December 2020. The list of ETFs includes Materials (XLB), Consumer Staples (XLP), Energy (XLE), Financial (XLF), Industrial (XLI), Technology (XLK), Utilities (XLU), Consumer Discretionary (XLY), and Health Care (XLV). The mean, standard deviation, expense ratio and bid-ask (BA) spread are presented as percentages. ARCH refers to the one-lag autocorrection coefficients for the squared returns. Leverage refers to the linear correlation coefficients between the lagged returns, r_{t-1} . and the squared return, r_t^2. The normality hypothesis is rejected for all assets using the Jarque-Bera test with 99% confidence. AUM refers to the assets under management in millions.

Table 4: Unconditional Pearson correlation and Ang & Chen (2002) exceedance correlation

	Median									
	Materials (XLB)	Consumer Staples (XLP)	Energy (XLE)	Financial (XLF)	Industrial (XLI)	Technology (XLK)	Utilities (XLU)	Consumer Discretionary (XLY)	HealthCare (XLV)	EWP
Materials (XLB)	1.00	0.60	0.71	0.74	0.84	0.73	0.50	0.75	0.66	0.88
Consumer Staples (XLP)	0.60	1.00	0.46	0.68	0.68	0.63	0.59	0.70	0.68	0.78
Energy (XLE)	0.71	0.46	1.00	0.61	0.68	0.60	0.49	0.60	0.54	0.78
Financial (XLF)	0.74	0.68	0.61	1.00	0.81	0.74	0.54	0.81	0.71	0.88
Industrial (XLI)	0.84	0.68	0.68	0.81	1.00	0.79	0.53	0.84	0.74	0.92
Technology (XLK)	0.73	0.63	0.60	0.74	0.79	1.00	0.45	0.80	0.69	0.86
Utilities (XLU)	0.50	0.59	0.49	0.54	0.53	0.45	1.00	0.49	0.48	0.68
Consumer Discretionary (XLY)	0.75	0.70	0.60	0.81	0.84	0.80	0.49	1.00	0.73	0.89
HealthCare (XLV)	0.66	0.68	0.54	0.71	0.74	0.69	0.48	0.73	1.00	0.83
EWP	0.88	0.78	0.78	0.88	0.92	0.86	0.68	0.89	0.83	1.00
25%-75%	0.05	0.00	0.04	0.05	0.03	-0.02	0.02	0.04	0.05	0.03

	IQR									
	Materials (XLB)	Consumer Staples (XLP)	Energy (XLE)	Financial (XLF)	Industrial (XLI)	Technology (XLK)	Utilities (XLU)	Consumer Discretionary (XLY)	HealthCare (XLV)	EWP
Materials (XLB)	0.00	0.14	0.26	0.13	0.07	0.15	0.35	0.11	0.17	0.05
Consumer Staples (XLP)	0.14	0.00	0.32	0.15	0.16	0.22	0.23	0.15	0.14	0.12
Energy (XLE)	0.26	0.32	0.00	0.32	0.33	0.40	0.30	0.38	0.30	0.22
Financial (XLF)	0.13	0.15	0.32	0.00	0.10	0.13	0.31	0.07	0.12	0.09
Industrial (XLI)	0.07	0.16	0.33	0.10	0.00	0.11	0.32	0.06	0.11	0.04
Technology (XLK)	0.15	0.22	0.40	0.13	0.11	0.00	0.29	0.13	0.15	0.09
Utilities (XLU)	0.35	0.23	0.30	0.31	0.32	0.29	0.00	0.36	0.33	0.28
Consumer Discretionary (XLY)	0.11	0.15	0.38	0.07	0.06	0.13	0.36	0.00	0.13	0.06
HealthCare (XLV)	0.17	0.14	0.30	0.12	0.11	0.15	0.33	0.13	0.00	0.11
EWP	0.05	0.12	0.22	0.09	0.04	0.09	0.28	0.06	0.11	0.00
25%-75%	0.22	0.26	0.25	0.18	0.14	0.13	0.26	0.19	0.20	0.20

This table presents the sample unconditional Pearson's correlations between the daily returns for the nine sectors ETFs and equally-weighted portfolio. The correlations are calculated for each estimation window, then medians and interquartile ranges (IQR) of the distributions are provided. The full sample runs from January 2000 to December 2020. The rows "25%-75%" show the difference between Ang & Chen (2002) exceedance correlation between the ETFs and the S&P 500 index, calculated for the 25th and 75th percentiles.

Table 5: Comparison of empirical historical density, risk-neutral and real-world densities based on the Heston pricing model

	EPD				RND				RWD				Maximum Likelihood			
	Mean	Std. Dev.	Skewness	Kurtosis	Mean	Std. Dev.	Skewness	Kurtosis	Mean	Std. Dev.	Skewness	Kurtosis	LL_1 (RND)	LL_2 (RWD)	LLR	p-val
S&P 500	0.78	4.44	-0.16	4.37	0.09	4.76	-1.28	5.92	0.41	4.02	-1.26	5.88	681.46	687.30	5.79	0.06
Materials (XLB)	0.94	5.42	-0.22	3.95	0.09	6.11	-0.78	4.78	0.34	4.85	-0.54	4.67	591.28	645.28	34.92	0.00
Consumer Staples (XLP)	0.76	3.53	-0.20	4.25	0.09	4.19	-0.93	6.20	0.40	3.34	-0.79	5.29	710.85	775.11	59.46	0.00
Energy (XLE)	0.70	7.09	-0.28	3.99	0.09	6.97	-0.40	3.64	0.10	5.92	-0.37	3.52	504.56	542.41	23.32	0.00
Financial (XLF)	0.74	5.38	-0.14	4.19	0.09	5.98	-0.86	4.96	0.57	4.70	-0.79	4.90	626.04	662.12	38.61	0.00
Industrial (XLI)	0.85	4.95	-0.16	4.16	0.09	5.45	-0.97	5.31	0.57	4.43	-0.78	4.87	654.33	683.74	34.89	0.00
Technology (XLK)	0.89	5.36	-0.16	4.44	0.09	5.75	-1.00	5.29	0.50	4.69	-0.96	5.30	613.88	642.49	22.56	0.00
Utilities (XLU)	0.90	4.38	-0.26	4.19	0.09	4.96	-0.64	4.77	0.24	3.81	-0.41	4.05	617.98	716.18	71.50	0.00
Consumer Discretionary (XLY)	0.99	5.03	-0.21	4.29	0.09	5.39	-1.03	5.47	0.55	4.46	-0.81	5.18	637.58	689.13	35.94	0.00
HealthCare (XLV)	0.69	4.30	-0.21	4.34	0.09	4.89	-0.95	5.73	0.48	3.91	-0.76	5.21	646.76	719.10	55.53	0.00
Median ETF	0.85	4.99	-0.21	4.19	0.09	5.45	-0.93	5.29	0.48	4.46	-0.78	4.90	626.04	683.74	35.94	0.00
EWP	0.92	4.18	-0.25	4.42	0.09	4.48	-0.66	4.32	0.92	3.40	-0.52	4.04	624.42	676.57	92.06	0.00
Top-1 EPD	1.92	5.67	-0.27	4.26	0.09	6.21	-0.80	4.86	0.43	5.08	-0.70	4.72	590.62	625.17	28.90	0.00
Top-1 RWD	0.91	4.83	-0.23	4.30	0.09	5.58	-0.90	4.86	3.59	2.80	-0.12	4.01	582.86	739.78	176.21	0.00

The table reports the median of the moments across 239 portfolio formation dates calculated using three probability distribution: empirical historical distribution based on 12 month estimation window (EPD); risk-neutral distribution based on options and Heston pricing model (RND); real-world distribution based on risk premium transformation of RND using 12 month estimation window. The results are provided for monthly return distributions. The last three columns report the loglikelihoods of RND (LL1) and RWD (LL2) with the log-likelihood test statistics ($LLR = 2(LL_2 - LL_1)$) and corresponding p-value with asymptotic chi-squared distribution with 2 degrees of freedom.

Table 6: Comparing density forecasts using the weighted likelihood ratio test

	Full distribution			Left tail of distribution		
	Heston RWD - RND	Heston RWD - BSM RWD	Heston RWD - GARCH	Heston RWD - RND	Heston RWD - BSM RWD	Heston RWD - GARCH
S&P 500	**11.56**	**2.41**	**7.88**	**10.77**	-0.60	**7.08**
Materials (XLB)	**2.28**	0.64	**7.67**	1.36	-0.49	**6.15**
Consumer Staples (XLP)	**1.69**	-0.70	**5.43**	0.99	**-1.99**	**5.10**
Energy (XLE)	0.91	**-1.73**	**5.94**	0.78	**-2.30**	**5.08**
Financial (XLF)	**2.21**	0.10	**7.58**	-0.43	**-3.18**	**4.55**
Industrial (XLI)	-0.02	0.78	**6.57**	**-1.81**	-1.57	**6.22**
Technology (XLK)	**3.51**	0.74	**7.73**	1.39	**-1.81**	**7.60**
Utilities (XLU)	**1.82**	0.74	**5.22**	1.45	0.59	**4.47**
Consumer Discretionary (XLY)	0.29	1.33	**6.13**	-0.74	-0.33	**5.57**
HealthCare (XLV)	**2.54**	**2.83**	**4.99**	1.61	1.38	**4.85**
Median ETF	**1.82**	0.74	**6.13**	0.99	-1.57	**5.10**
EWP	**3.85**	1.46	**6.57**	1.45	**-1.98**	**5.69**
Top-1 EPD	0.06	-1.27	**5.96**	-1.19	-1.64	**5.49**
Top-1 RWD	**2.24**	**1.65**	**5.45**	0.14	-0.84	**3.52**

This table reports the results for the weighted likelihood ratio (WLR) test by Amisano and Giacomini (2007). The null hypothesis is that the two comparing models f and g perform identically in terms of density forecasting, The WLR is calculated as follows:

$$WLR_{t+1} = w\left(P_{t+1}\right)\left(\log\left(\hat{f}_t\left(P_{t+1}\right)\right) - \log\left(\hat{g}_t\left(P_{t+1}\right)\right)\right), t = 1, \ldots, T-1,$$

where P_{t+1} is the price at time $t+1$, f_t and g_t are density forecasts estimated at time t using the model f and g, $w()$ is the weight function, $T-1$ is the number of forecasting periods. The test is conducted for 9 sectors ETFs and S&P500 index. The out-of-sample period is from January 2001 to December 2020, where $T = 240$ months. The table reports the test statistics $t = \bar{WLR}/\hat{\sigma}\sqrt{T-1}$, where $\hat{\sigma}$ is HAC estimator of the WLR asymptotic variance. The t-statistic is normally distributed with mean 0 and variance 1. The first three columns report results with no weighting ($w() = 1$). The last three columns report results for the left-tail-weighted case where the tails of distribution have higher left weights $w()$ values than the upper part of distribution. The header of each column indicates the comparing models' names ($f - g$). Positive values indicate that f outperforms g and vice versa. The bold values indicate that the null is reject for at least 10 % significance level.

Table 7: The copula fit analysis

	Dependence measures								Model fitting					
	Median				IQR				Loglikelihood		Akaike		Schwartz	
	Gaussian		RVine		Gaussian		RVine							
	Corr	25%-75%	Corr	25%-75%	Corr	25%-75%	Corr	25%-75%	Gaussian	RVine	Gaussian	RVine	Gaussian	RVine
Materials (XLB)	0.63	0.00	0.61	0.05	0.14	0.00	0.12	0.10	27.57	29.45	-53.14	-56.47	-49.61	-52.19
Consumer Staples (XLP)	0.58	0.00	0.57	0.03	0.16	0.00	0.15	0.08	18.83	20.43	-35.67	-38.51	-32.14	-34.37
Energy (XLE)	0.50	0.00	0.48	0.06	0.31	0.00	0.25	0.11	14.16	15.54	-26.31	-28.94	-22.79	-25.16
Financial (XLF)	0.64	0.00	0.63	0.06	0.11	0.00	0.11	0.12	26.91	28.83	-51.81	-55.22	-48.29	-50.90
Industrial (XLI)	0.68	0.00	0.67	0.05	0.13	0.00	0.12	0.08	58.23	61.31	-114.46	-119.75	-110.93	-114.69
Technology (XLK)	0.63	0.00	0.59	0.04	0.13	0.00	0.12	0.08	22.44	24.49	-42.89	-46.52	-39.36	-42.19
Utilities (XLU)	0.44	0.00	0.42	0.02	0.27	0.00	0.21	0.09	9.44	10.65	-16.89	-19.12	-13.36	-15.27
Consumer Discretionary (XLY)	0.65	0.00	0.65	0.04	0.13	0.00	0.11	0.09	38.83	40.88	-75.65	-79.18	-72.13	-74.62
HealthCare (XLV)	0.59	0.00	0.57	0.04	0.13	0.00	0.13	0.10	18.33	20.10	-34.66	-37.87	-31.13	-33.73
Median	0.63	0.00	0.59	0.04	0.13	0.00	0.12	0.09	22.44	24.49	-42.89	-46.52	-39.36	-42.19
EWP	0.59	0.00	0.57	0.05	0.15	0.00	0.14	0.08	26.08	27.97	-50.16	-53.51	-46.64	-49.24
Top-1 EPD	0.59	0.00	0.56	0.04	0.26	0.00	0.25	0.10	20.85	22.57	-39.69	-42.77	-36.17	-38.59
Top-1 RWD	0.63	0.00	0.61	0.04	0.16	0.00	0.17	0.10	31.31	33.29	-60.62	-64.11	-57.09	-59.78

The table reports the median and interquartile ranges (IQR) of several dependence and model fit measures implied by Gaussian multivariate copula and Regular Vine (RVine) multivariate copula with various pair-copula types. The dependence measures are unconditional linear correlation (Corr) and the difference between Ang and Chen (2002) exceedance correlation, calculated for the 25th and 75th percentiles ("25%-75%" show). The model fit measures are Log-Likelihood, Akaike and Schwarz (Bayesian) information criterion. For each of the 9 ETFs, the dependence and model fit measures are averaged across the 8 fund pairs that include the relevant fund.

Table 8: The out-of-sample performance (base case)

	Passive		Historical			Option-implied					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	SPX	EWP	Top-1	Top-3	GARCH	BSM	Heston	RVine	Wins	MVD	Heston/RND
Mean	0.627	0.831	0.871	0.800	0.748	0.761	1.226	1.157	1.232	1.024	0.676
TW	311.82	505.24	512.46	487.26	442.56	429.67	1242.54	1046.47	1297.76	788.53	388.74
Stand. Dev.	5.32	5.33	6.04	5.10	4.86	5.38	5.64	5.69	5.49	5.45	4.46
Semidev.	4.00	3.95	4.04	3.69	3.51	3.80	3.86	3.93	3.59	3.83	3.27
Skewness	-1.43	-1.59	-0.38	-1.16	-1.28	-0.77	-1.10	-1.15	-0.55	-1.24	-1.74
VaR95%	8.20	7.86	9.49	7.17	7.02	8.34	6.79	7.20	6.79	6.76	5.04
CVaR95%	14.53	14.17	13.75	13.20	12.24	13.63	13.50	13.56	12.62	13.13	11.17
CE	-0.06	0.12	0.08	0.20	0.20	0.11	0.46	0.37	0.56	0.30	0.17
Sharpe	0.10	0.14	0.13	0.14	0.13	0.12	0.20	0.18	0.20	0.17	0.13
Sortino	0.13	0.18	0.19	0.19	0.18	0.17	0.29	0.27	0.31	0.24	0.17
Turnover		0.02	0.69	0.46	1.04	0.87	1.37	1.35	1.37	1.35	0.54
Concentration		0.11	1.00	0.33	0.56	0.68	0.86	0.87	0.92	0.76	0.40
Mean t-test		**0.000**	0.152	0.123	0.201	0.240	**0.000**	**0.002**	**0.001**	**0.006**	0.388
LW08 p-value		**0.004**	0.507	0.145	0.254	0.495	**0.005**	**0.021**	**0.002**	**0.030**	0.406
LMW05 p-value		0.210	0.152	0.843	0.719	0.743	0.581	0.586	0.714	0.590	0.343

This table presents the out-of-sample performance of competing portfolio strategies over the sample period from January 2001 to December 2020. Rebalancing happens on the first trading day of every month. The performance measures are estimated from monthly returns. The measures shown are the mean, standard deviation, semi-deviation, and skewness, as well as the historical value at risk (VaR) and conditional value at risk (CVaR) for the 95% confidence level, the certainty equivalent (CE) return for a power utility function with RRA = 4, the Sharpe ratio, the Sortino ratio, turnover and the sum of squared portfolio weights (Concentration). Mean t-test displays p-values for testing whether the average return spread between the corresponding portfolio and the benchmark S&P 500 equals zero. The LW08 p-value is a p-value for the Sharpe ratio difference test between the corresponding portfolio and the benchmark S&P 500, estimated using robust bootstrap inference, as suggested by Ledoit and Wolf (2008). The LMW05 p-value is a p-value for the Linton et al. (2005) SSD statistical tests. The bold p-values indicate that the null is rejected for at least 10% significance level. For excess returns, we use 0.11% as the monthly risk-free rate.

Table 9: The out-of-sample performance net of transaction costs

	Passive		Historical			Option-implied					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	SPX	EWP	Top-1	Top-3	GARCH	BSM	Heston	RVine	Wins	MVD	Heston RND
Mean	0.627	0.826	0.758	0.730	0.595	0.643	1.057	0.992	1.062	0.851	0.601
TW	311.82	500.38	393.12	412.54	314.18	340.55	861.72	735.11	879.61	539.43	328.03
Stand. Dev.	5.32	5.34	6.02	5.10	4.86	5.40	5.63	5.69	5.48	5.45	4.47
Semidev.	4.00	3.96	4.09	3.72	3.60	3.87	3.93	4.00	3.67	3.90	3.32
Skewness	-1.43	-1.59	-0.39	-1.17	-1.34	-0.76	-1.15	-1.19	-0.63	-1.25	-1.76
VaR95%	8.20	7.87	9.49	7.24	7.02	8.40	6.80	7.24	6.80	7.33	5.34
CVaR95%	14.53	14.21	13.86	13.27	12.59	13.73	13.66	13.72	12.76	13.32	11.43
CE	-0.06	0.11	-0.02	0.13	0.04	-0.01	0.29	0.20	0.39	0.13	0.09
Sharpe	0.10	0.13	0.11	0.12	0.10	0.10	0.17	0.15	0.17	0.14	0.11
Sortino	0.13	0.18	0.16	0.17	0.13	0.14	0.24	0.22	0.26	0.19	0.15
Turnover		0.02	0.71	0.47	1.04	0.87	1.37	1.36	1.38	1.35	0.54
Concentration		0.11	1.00	0.33	0.56	0.68	0.86	0.87	0.92	0.75	0.40
Mean t-test		**0.000**	0.312	0.244	0.628	0.498	**0.010**	**0.029**	**0.011**	**0.093**	0.593
LW08 p-value		**0.003**	0.808	0.343	0.993	0.972	**0.041**	0.129	**0.029**	0.251	0.769
LMW05 p-value		0.211	**0.081**	0.708	0.455	0.340	0.593	0.598	0.718	0.656	0.292

This table presents the out-of-sample performance calculated net of proportional transaction costs. Rebalancing happens on the first trading day of every month. The performance measures are estimated from monthly returns. The measures shown are the mean, terminal wealth(TW), standard deviation, semi-deviation, and skewness, as well as the historical value at risk (VaR) and conditional value at risk (CVaR) for the 95% confidence level, the certainty equivalent (CE) return for a power utility function with RRA = 4, the Sharpe ratio, the Sortino ratio, turnover and the sum of squared portfolio weights (Concentration). We model TW by hypothetically investing \$100 at the start of the out-of-sample periods for each portfolio. Mean t-test displays p-values for testing whether the average return spread between the corresponding portfolio and the benchmark S&P 500 equals zero. The LW08 p-value is a p-value for the Sharpe ratio difference test between the corresponding portfolio and the benchmark S&P 500, estimated using robust bootstrap inference, as suggested by Ledoit and Wolf (2008). The LMW05 p-value is a p-value for the Linton et al. (2005) SSD statistical tests. The bold p-values indicate that the null is rejected for at least 10% significance level. For excess returns, we use 0.11% as the monthly risk-free rate.

Table 10: The out-of-sample performance in periods of high and low VIX (volatility)

Panel A. High VIX											
	Passive		Historical			Option-implied					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	SPX	EWP	Top-1	Top-3	GARCH	BSM	Heston	RVine	Wins	MVD	Heston RND
Mean	-0.620	-0.412	-0.344	-0.649	-0.318	-0.551	0.524	0.252	0.564	0.281	-0.404
TW	53.06	59.68	64.62	54.78	67.52	56.79	103.62	87.79	109.83	91.21	64.73
Stand. Dev.	9.03	9.14	8.63	8.19	8.00	8.69	9.40	9.42	8.95	9.05	7.73
Semidev.	7.26	7.28	6.52	6.63	6.28	6.73	6.96	7.09	6.31	6.86	6.11
Skewness	-0.75	-0.85	-0.31	-0.66	-0.77	-0.35	-0.75	-0.75	-0.29	-0.85	-0.94
VaR95%	16.47	15.46	13.54	16.10	15.47	17.88	11.70	11.10	11.30	11.11	12.17
CVaR95%	24.83	25.61	20.88	22.78	20.36	21.74	27.37	27.82	24.07	27.03	22.58
CE	-2.57	-2.48	-1.98	-2.19	-1.81	-2.21	-1.63	-1.94	-1.23	-1.74	-1.92
Sharpe	-0.08	-0.06	-0.05	-0.09	-0.05	-0.08	0.04	0.02	0.05	0.02	-0.07
Sortino	-0.10	-0.07	-0.07	-0.11	-0.07	-0.10	0.06	0.02	0.07	0.02	-0.08
Concentration		0.11	1.00	0.33	0.56	0.68	0.86	0.87	0.92	0.76	0.40
Mean t-test		0.086	0.314	0.531	0.214	0.447	**0.012**	**0.051**	**0.014**	**0.021**	0.334
Variance t-test		0.851	0.674	0.331	0.227	0.730	0.615	0.596	0.972	0.930	0.122
LMW05 p-value		0.506	0.839	0.935	0.258	0.620	0.483	0.506	0.406	0.517	0.657
Panel B. Low VIX											
	Passive		Historical			Option-implied					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	SPX	EWP	Top-1	Top-3	GARCH	BSM	Heston	RVine	Wins	MVD	Heston RND
Mean	1.599	1.719	1.805	1.553	1.545	1.812	1.765	1.812	1.677	1.479	1.567
TW	255.89	274.82	272.28	246.17	248.11	285.21	279.43	286.99	265.72	236.46	251.80
Stand. Dev.	2.04	2.03	5.01	2.86	1.98	3.24	2.77	2.86	2.68	2.64	1.81
Semidev.	0.84	0.83	2.70	1.40	0.76	1.40	1.24	1.25	1.27	1.42	0.63
Skewness	-0.42	-0.54	0.05	-0.37	-0.37	0.77	-0.24	-0.18	-0.45	-0.81	-0.32
VaR95%	2.05	2.44	6.03	2.60	2.20	2.54	3.13	3.13	3.13	3.40	1.81
CVaR95%	3.27	3.24	10.50	5.23	2.76	5.61	4.55	4.52	4.72	5.54	2.31
CE	1.52	1.64	1.31	1.39	1.47	1.61	1.61	1.65	1.53	1.34	1.50
Sharpe	0.73	0.79	0.34	0.50	0.72	0.53	0.60	0.59	0.58	0.52	0.80
Sortino	1.76	1.94	0.63	1.03	1.88	1.22	1.33	1.36	1.23	0.96	2.29
Concentration		0.11	1.00	0.33	0.56	0.68	0.86	0.87	0.92	0.76	0.40
Mean t-test		**0.022**	0.367	0.566	0.620	0.275	0.279	0.230	0.385	0.690	0.560
LW08 p-value		0.138	**0.039**	**0.061**	0.951	0.138	0.318	0.289	0.269	**0.072**	0.575
LMW05 p-value		0.903	**0.000**	0.258	0.871	0.774	**0.097**	**0.097**	**0.097**	**0.097**	0.323

The table shows the out-of-sample performances during periods of high and low market volatility proxied by the VIX index. Panel A shows the results for the period of high VIX, which is identified as the months when the VIX is above the 75th percentile of the VIX values distribution. Panel A shows the results for the period of low VIX, which is identified as the months when the VIX is below the 25th percentile of the VIX values distribution.

Appendix A. Portfolio optimization based on Stochastic Dominance

To apply stochastic dominance to portfolio optimization, we introduce some designations here. Assume that the investment universe consists of N assets with random investment returns, $\boldsymbol{R} = (R_1, R_2, \ldots, R_N)$, whose support is bounded by $[R_{min}, R_{max}]^N$, $-100\% < R_{min} < R_{max} < \infty$. The investment opportunity set consists of all convex combinations of the assets, $\Omega := \left\{\boldsymbol{\omega} \in \mathbb{R}_+^N : 1_N^{\mathrm{T}}\boldsymbol{\omega} = 1; \boldsymbol{\omega} \geq 0\right\}$, allowing for all long-only portfolios. The optimization objective is to find a portfolio which will dominate the prespecified benchmark, $\tau \in \Omega$, with returns equal to $y := R^T\boldsymbol{\tau}$.

Let $F(\boldsymbol{R}) : \mathbb{R}^N \Rightarrow [0,1]$ denote the joint cumulative distribution function (cdf) of R and

$$F_\omega(z) = \int_{\left\{\boldsymbol{R} \in R^N; \boldsymbol{R}^{\mathrm{T}}\boldsymbol{\omega} \leq z\right\}} dF(\boldsymbol{R}) \tag{A.1}$$

be the marginal cdf for portfolio $\omega \in \Omega$. The cdf is latent and needs to be estimated using one of the methods shown in the sections that will follow. To implement the Second-degree SD concept (Hadar and Russell, 1969; Hanoch and Levy, 1969; Rothschild and Stiglitz, 1970), Bawa (1975) introduces the first-order lower-partial moment (lpm):

$$\mathcal{L}_\omega(z) := \int_{R_{\min}}^{z} F_\omega(R)dR = E_{F_\omega}[(z-R)I(R \leq z)], \tag{A.2}$$

where $I(r \leq z)$ is an indicator function that takes on a value of unity if the condition within the parentheses is satisfied, and zero otherwise. Portfolio $\boldsymbol{\omega}$ then dominates the benchmark, $\boldsymbol{\tau}$, by SSD if and only if:

$$\mathcal{L}_\omega \leq \gamma_\tau, \forall z \in [R_{\min}, R_{\max}]. \tag{A.3}$$

To allow for a numerical analysis, applications of SSD generally use a discrete cdf estimator $\hat{F}$ with atoms R_t and probabilities p_t, $t = 1, \ldots, T$. In our application, continuous cdf estimates are discretized using Monte Carlo simulation methods. Each return vector R_t in the random samples is equiprobable so that $p_t = 1/T$. Using the discrete cdf estimator, the lpm can be estimated as follows:

$$\hat{\gamma}_\omega(z) = \sum_{t=1}^{T} p_t \left(z - \boldsymbol{R}_t^{\mathrm{T}}\boldsymbol{\omega}\right) I\left(\boldsymbol{R}_t^{\mathrm{T}}\boldsymbol{\omega} \leq z\right). \tag{A.4}$$

Since the estimated lpm takes on a simple piecewise-linear, increasing, and convex shape, checking the dominance condition only at the observed return levels for the benchmark ($z = y_t, t = 1, \ldots, T$) would suffice. The lpm can be linearized using the lifting representation by Rockafellar and Uryasev (2000).

Combining the SSD conditions with a budget constraint and short-selling constraints in Ω, we obtain SSD enhanced portfolios as solutions to the following system of constraints:

$$\begin{aligned} \hat{\gamma}_{\omega}(y_i) \leq \hat{\gamma}_{\tau}(y_i), \quad i = 1, \ldots, T, \\ \mathbf{1}_N^{\mathrm{T}} \boldsymbol{\omega} = 1, \\ \omega_j \geq 0, \quad j = 1, \ldots, M. \end{aligned} \tag{A.5}$$

Any feasible solution, $\boldsymbol{\omega}*$, to this system dominates the benchmark portfolio, $\boldsymbol{\tau}$, by SSD.

The objective function for portfolio optimization using SSD is then found as follows:

$$G_{\hat{F},\omega} := E_{\hat{F}_{\omega}}[R] - \sum_{t=1}^{T} \lambda_t \hat{\gamma}_{\omega}(y_t), \tag{A.6}$$

where $\lambda_t \leq 0, t = 1, \ldots, T$ are decision weights. In the empirical application, the objective function is the average portfolio return. This common specification amounts to $\lambda_t = 0$, $t = 1, ..., T$. This objective of the SSD optimization problem can theoretically lead to a solution which is not fully efficient. Secondary performance improvements can be achieved using multi-stage programming. The effect of this method on out-of-sample performance is, however, negligible in this study.

The problem of maximizing this objective subject to the system of constraints is a well-known convex optimization problem which can be solved efficiently using the linearization of the lpms by Rockafellar and Uryasev (2000) and/or cutting plane algorithms as in Fábián et al. (2011). The present study uses the Aorda Portfolio Safeguard Toolbox for MATLAB.[1]

The portfolio optimization model requires an estimate for the joint probability distribution of the base assets. In this section, we describe an estimation procedure using

[1] http://www.aorda.com/index.php/portfolio-safeguard/

options data. The proposed procedure consists of three parts: 1) an estimation of risk-neutral process and distribution, 2) a transformation of the risk-neutral to the real-world probability distribution, and 3) the joint distribution function estimation. For the sake of comparison, the option-implied estimates are compared with historical estimates for the marginal distributions which are based on GARCH.

Appendix B. Risk-neutral process

To estimate the option-implied distribution, we assume two alternative specifications for the risk-neutral stochastic process of the underlying assets. The simple geometric Brownian motion (GBM) is the basis for the BSM option pricing formula:

$$\frac{dS}{S} = (r-q)dt + \sqrt{V}dW. \tag{B.1}$$

Here, r is the risk-free rate, q is the dividend yield, V is the variance and W is a Wiener process. The risk-neutral distribution of $\log(S_T)$ at time T is then normal:

$$\log(S_T) \sim N\left(\log(S_0) + (r-q)T - 0.5VT, VT\right). \tag{B.2}$$

The risk-neutral density of S_T then depends on the parameters, (r, q, S_0, V, T), and is given by the lognormal density

$$f_{Q,T}(S) = \frac{1}{S_T\sqrt{2V\pi T}} exp(-\frac{1}{2}\frac{(log(S_T) - [log(S_0) + (r - q - \frac{1}{2}V^2)T])^2}{VT}). \tag{B.3}$$

A more general model assumes volatility is stochastic. Heston (1993) stochastic volatility process is chosen because it has a closed-form characteristic function and theoretical option prices, whose implied volatilities show realistic term structure and smile effects. Shackleton et al. (2010) found that the option-implied probabilities estimated from the Heston pricing model were more informative than historical distributions or option-implied probabilities estimated using the BSM model for a forecast horizon from 2 to 12 weeks. The risk-neutral process with stochastic volatility is described by the following set of equations:

$$\begin{aligned} dS/S &= (r-q)dt + \sqrt{V}dW_1, \\ dV &= \kappa(\theta - V)dt + \sigma\sqrt{V}dW_2, \\ \text{cor}\,(W_1, W_2) &= \rho. \end{aligned} \tag{B.4}$$

Here, θ is the level towards which the stochastic variance, V, reverts and κ denotes the rate of reversion of V toward θ. σ is the volatility of the volatility parameter, while ρ is the correlation between the two Wiener processes, W_1 and W_2. Relaxation of the constant volatility assumption allows nonnormality of the underlying distribution, which is more realistic, given previous empirical studies. In particular, the correlation, ρ, controls the skewness of the distribution, while the volatility of volatility, σ, affects the kurtosis.

Heston (1993) provides a formula for pricing European style call options based on the above process:

$$C(S_0, K) = S_0 e^{-qT} P_1 - K e^{-rT} P_2. \tag{B.5}$$

Here, P_j, $j = 1, 2$ are the conditional probabilities that the option expires in the money. P_2 is a risk-neutral probability, while P_1 is a probability under another measure with a different drift rate. The risk-neutral measure Q uses the riskless asset as numeraire, while the alternative measure for P_1 uses the stock price S_0.[2] We use this theoretical price formula to estimate the risk-neutral parameters, by matching theoretical and market option prices for all expiration dates and moneyness levels.

Given the parameter values, the risk-neutral distribution for different horizons can be derived by numerical integration of the characteristic functions of models. Heston's (1993) characteristic functions have a risk-neutral expectation form, for any real number:

$$g(\psi) = E_Q\left[\exp\left(i\psi \log\left(S_T\right)\right)\right], \tag{B.6}$$

where $i = \sqrt{-1}$. Using an inversion formula, the risk-neutral density, $f_{Q,T}$, can then be estimated:

$$f_{Q,T}(S) = \frac{1}{\pi S} \int_0^\infty \text{Real}[\exp(-i\psi \log(S)) g(\psi)] d\psi, \tag{B.7}$$

where we use only the real part of the complex integrand. The latent parameters are

[2] The relation between P_1 and P_2 is following:

$$P_1 = E_Q\left[\frac{dQ^S}{dQ} I\left(S_T > K\right)\right]; P_2 = E_Q\left[I\left(S_T > K\right)\right]; \frac{dQ^S}{dQ} = \frac{S_T}{S_0 e^{rT}}$$

where $E_Q[...]$ is a risk-neutral expectation, $I(...)$ is an indicator function. The last term is a Radon-Nikodym derivative of alternative measure Q^S with respect to risk-neutral measure Q.

estimated monthly on the last trading day (day before the rebalancing date) from the whole cross-section of options contracts. The estimated volatility of the GBM process is provided by the simplest credible estimate, namely the end-of-day, nearest-the-money implied volatility for the nearest-to-expiry options.

On day t, for the Heston model, we estimate the initial variance, V_t, the three volatility parameters, θ_t, κ_t, and σ_t, and the correlation, ρ_t. Suppose N_t European call option contracts are traded on day t, indexed by $i = 1, ..., N_t$, with strikes $K_{t,i}$, expiry times $T_{t,i}$, and market prices $c_{t,i}$; additionally, suppose $p_{t,i} = S_0 \exp(r - q)T$ is the futures price for the equity, calculated for a synthetic futures contract that expires after $T_{t,i}$ years.[3] The five Heston parameters are then estimated by minimising:

$$\sum_{i=1}^{N_t} \left(c_{t,i} - c\left(p_{t,i}, K_{t,i}, T_{t,i}, V_t, \kappa_t, \theta_t, \sigma_t, \rho_t\right)\right)^2, \tag{B.8}$$

where $c(...)$ is the Heston pricing formula in Eq. (B.5). Christoffersen and Jacobs (2004a) conclude that the mean squared pricing error is a '... good general-purpose loss function in option valuation applications ...'. Subsequently, it has also been employed in studies on equity dynamics, notably by Christoffersen et al. (2010), Shackleton et al. (2010), and Fan et al. (2018).

Appendix C. Transformation methods

We apply two established transformation methods here based on Liu et al. (2007) and Shackleton et al. (2010). For the lognormal distribution of the BSM, we use a statistical calibration with a two-parameter beta distribution as a calibration function. For the Heston model, we apply an alternative, more analytically tractable risk transformation approach based on the direct adjustment of the risk premia.

[3]Following Shackleton et al. (2010), we assume that the same price dynamics and parameter values apply to the set of contemporaneous futures contracts. Fan et al. (2018) prove that this assumption is reasonable.

Appendix C1. Statistical calibration.

The calibration method relies on learning from past outcomes how best to change a measure from risk-neutral to real-world. It is well known that the cumulative probabilities are uniformly distributed for correctly specified densities. Observed deviations from uniformity can be exploited to transform risk-neutral densities and to obtain better descriptions of real-world outcomes.

At a starting point, $t = 0$, we forecast the risk-neutral density and cdf from a cross-section of options and denote them by $f_{Q,H}$ and $F_{Q,H}$, respectively. Let the calibration function, $C_H(u)$, be the real world cdf of a random variable, $U_H = F_{Q,H}(S_H)$, that depends on the forecast horizon, H. The real world cdf, $F_{P,H}$, then equals the calibrated risk-neutral cdf (Liu et al., 2007): $F_{P,H}(S) = C_H(F_{Q,H}(S))$. Taking the derivative of $F_{P,H}$, we obtain the real-world density, $f_{P,H} = f_{Q,H}(S)c_T(u)$, where $u = F_{Q,H}(S_H)$ and $c_H(u)$ is the real-world density of U_H.

The parametric calibration function, $C_H(u)$, is the cdf of the Beta distribution suggested by Fackler and King (1990) and later applied by Liu et al. (2007) and Shackleton et al. (2010). The calibration density is then:

$$c_T = u^{a-1}(1-u)^{b-1}/B(a,b), 0 \leq u \leq 1, \tag{C1.1}$$

where $B(a, b)$ is a beta function. If the calibration parameters $a = b = 1$, then the investor is risk-neutral and $f_{Q,H} = f_{P,H}$. Given n observations, the parameters for the next period, $(n + 1)$, can be estimated using the maximum likelihood method.

Appendix C2. Atiya and Wall (2009) method

For the Heston model, the application of statistical beta calibration is complicated by the absence of analytical cdf and inverse cdf. Therefore, we apply an alternative method based on the adjustment of risk premia. Shackleton et al. (2010) show that this approach yields similar results as parametric calibration when both the price and the variance equations are transformed. Therefore, this approach is suitable for the Heston bivariate model. Simulation under the risk premium transformation does not require inverse cdf calculations and is based on a standard discretisation scheme such as Milstein one (Kahl and Jäckel, 2006).

The Atiya and Wall (2009) method for risk-premium transformation requires us to specify the bivariate real-world stochastic process: the logarithmic adjusted price, $y = \log(S)$, and the variance, V:

$$\begin{aligned} dy &= (r + \mu V)\, dt + \sqrt{V} d\tilde{W}_1, \\ dV &= (\alpha - \beta V) dt + \sigma \sqrt{V} d, \tilde{W}_2 \\ \operatorname{cor}\left(\tilde{W}_1, \tilde{W}_2\right) &= \rho, \end{aligned} \tag{C2.1}$$

where r is the risk-free rate; $\mu = \lambda_1(1-\rho^2) + \lambda_2\rho - 1/2$, $\alpha = \kappa\theta$ and $\beta = (\kappa - \lambda_2\sigma)$, where λ_1 and λ_2 are market prices of equity and stochastic volatility risks, θ is the level towards which the stochastic variance, V, reverts, κ denotes the rate of reversion of V towards θ, σ is the volatility of of stochastic variance, ρ is the correlation between the two Brownian motions. The risk-neutral parameters $\{V_t,\, \kappa_t,\, \theta_t,\, \sigma_t,\, \rho_t\}$ are estimated by minimizing Eq. (B.8). The risk premium parameters $\{\lambda_1,\, \lambda_2\}$ are estimated using historical daily returns data.

Since the Brownian motions are correlated normal random variables, the transition probability density for the joint log-stock price/variance process from time t to $t+1$ is bivariate normal:

$$p\left(y_{t+1}, V_{t+1} \mid y_t, V_t\right) = \Phi\left(\boldsymbol{M}_{t+1}, \boldsymbol{\Sigma}_{t+1}\right), \tag{C2.2}$$

where $\Phi\left(\boldsymbol{M}_{t+1}, \boldsymbol{\Sigma}_{t+1}\right)$ is the bivariate normal density with mean vector

$$\boldsymbol{M}_{t+1} = \begin{bmatrix} y_t + (r + \mu V_t)\, dt \\ V_t + (\alpha + \beta V_t)\, dt \end{bmatrix} \tag{C2.3}$$

and covariance matrix

$$\Sigma_{t+1} = V_t dt \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix}, \tag{C2.4}$$

where dt is the time increment between t and $t+1$, $t = 0, ..., n-1$. Since the Atiya and Wall (2009) method approximates a continuous process with discrete observations, the size of dt defines the precision of the method. In the case of daily frequency, $dt \approx 1/251$. Atiya and Wall (2009) apply a filtering argument to show that the likelihood for the unobserved variances can be approximated from the likelihood of stock prices. The likelihood at time $t+1$, given a value, V_t, at time t, is proportional to:

$$L_{t+1}\left(V_{t+1}\right) \propto d_t \left(ab_t\right)^{-\frac{1}{4}} e^{-2\sqrt{ab_t}} L_t\left(\sqrt{\frac{b_t}{a}}\right). \tag{C2.5}$$

This expression uses the following inputs:

$$\begin{aligned} a &= \frac{\beta'^2 - 2\mu\rho\sigma\beta' dt - \mu^2\sigma^2 dt^2}{2\sigma^2\left(1-\rho^2\right)dt} \\ b_t &= \frac{\left(V_{t+1} - \alpha dt\right)^2 - 2\rho\sigma\left(V_{t+1} - \alpha dt\right)\left(dy_{t+1} - rdt\right) + \sigma^2\left(dy_{t+1} - rdt\right)^2}{2\sigma^2\left(1-\rho^2\right)dt} \\ d_t &= \frac{\exp\left[\frac{\left(2\beta' - 2\mu\rho\sigma dt\right)\left(V_{t+1} - \alpha dt\right) - \left(2\rho\sigma\beta' - 2\mu\sigma^2 dt\right)\left(dy_{t+1} - rdt\right)}{2\sigma^2\left(1-\rho^2\right)dt}\right]}{D} \\ \beta' &= 1 - \beta dt \\ D &= 2\pi\sigma\sqrt{1-\rho^2}\,dt \\ dy_{t+1} &= y_{t+1} - y_t \end{aligned} \tag{C2.6}$$

The likelihood function depends also on V_{t+1}, through the terms b_t and d_t. To evaluate V_{t+1} from V_t, Atiya and Wall (2009) note that $V_t = \sqrt{b_t/a}$. Inverting this expression and applying the quadratic formula produces the solution

$$V_{t+1} = \sqrt{B^2 - C} - B, \tag{C2.7}$$

where

$$B = -\alpha dt - \rho\sigma\left(dy_{t+1} - rdt\right),$$

$$C = \left(\alpha dt\right)^2 + 2\rho\sigma\alpha dt\left(dy_{t+1} - rdt\right) + \sigma^2\left(dy_{t+1} - rdt\right)^2 - 2V_t^2 a\sigma^2\left(1-\rho^2\right)dt.$$

Hence, to evaluate the likelihood Eq. (C2.5), we start with initial values $L_0(V_0)$. To evaluate $L_{t+1}(V_{t+1})$ given $L_t(V_t)$, we first obtain V_{t+1} from Eq. (C2.7). We then apply Eq. (C2.5). We continue until the last observation n, at which time we have $L_T(V_n)$, to which we apply an optimization routine to find the maximum:

$$[\lambda_1, \lambda_2] = \underset{\lambda_1, \lambda_2}{\operatorname{argmax}} L_n(V_n). \tag{C2.8}$$

The initial value suggested by Atiya and Wall (2009) is $L_0(V_0) = \exp(-V_0)$, where V_0 is the initial risk-neutral variance parameter from the Heston model.

Appendix D. Joint probability distribution

According to Sklar's theorem, the joint probability distribution of $F(\boldsymbol{R})$, where $\boldsymbol{R}$ is a vector $(R_1, R_2, \ldots, R_N)$, can be represented using marginal probability functions, $F_i(R_i), i = 1, 2, \ldots, N$, and copula (dependence) function C:

$$F(\boldsymbol{R}) = C\left(F_1\left(R_1\right), F_2\left(R_2\right), \ldots, F_N\left(R_N\right)\right), \tag{D.1}$$

where a copula is also a joint distribution $C(u)$ with marginals $u_i, i = 1, 2, \ldots, N$, that are uniform on the unit interval [0,1]. That is, we can describe the joint distribution of $\boldsymbol{R}$ by the marginal distributions $F_i(R_i)$ and the copula C. From a modeling perspective, Sklar's Theorem allows us to separate the modeling of the marginal distributions $F_i(R_i)$ from the dependence structure.

Taking the partial derivative of the cdf in Eq. (D.1) with respect to each variable, $R_i, i = 1, 2, \ldots, N$, we obtain:

$$f(\boldsymbol{R}) = \prod_{i=1}^{N} f_i\left(R_i\right) c\left(F_1\left(R_1\right), F_2\left(R_2\right), \ldots, F_n\left(R_n\right)\right) \tag{D.2}$$

Here, $f(\boldsymbol{R})$ is the joint probability density function (pdf), $f_i(R_i)$ is the marginal pdf, and c is the copula density function. The standard Gaussian implies that the Pearson linear correlation matrix fully describes the dependence structure between variables. However, many studies reveal that the equity returns exhibit lower tail dependence (Ang and Chen, 2002; Patton, 2004; Low et al., 2013). In this case, Gaussian copula can be a poor approximation of the real dependence between equities.

The regular vine (R-Vine) copula is a flexible dependence model which consists of conditional pair-copula blocks. It enables specifying various types of dependence (including left tail asymmetry) by choosing appropriate bivariate copula families for each pair of variables. In the case of R-Vine copula, the Eq. (D.2) can be decomposed into a product of bivariate copulas:

$$f(\boldsymbol{R}) = \prod_{i=1}^{N} f_i \prod_{k=N-1}^{1} \prod_{i=N}^{k+1} c_{m_{k,k},m_{i,k}|m_{i+1,k},\ldots,m_{N,k}} \left(F_{m_{k,k}|m_{i+1,k},\ldots,m_{N,k}}, F_{m_{i,k}|m_{i+1,k},\ldots,m_{N,k}}\right), \tag{D.3}$$

where the arguments of all the functions have been omitted to shorten the notation. The dependence structure and conditional bivariate copulas, $c_{m_{k,k},m_{i,k}|m_{i+1,k},\ldots,m_{N,k}}$, are described using matrix notation, where indexation of $c_{m_{k,k},m_{i,k}|m_{i+1,k},\ldots,m_{N,k}}$ corresponds to the elements of an R-Vine lower triangular M matrix, where diagonal elements are in a descending order, $m_{k,k} = N - k + 1$ (Dissmann et al., 2013):

$$M = \begin{bmatrix} m_{1,1} & 0 & 0 \\ \vdots & \ddots & 0 \\ m_{1,N} & \cdots & m_{N,N} \end{bmatrix}. \tag{D.4}$$

The R-Vine copula approach is widely used in financial studies (see Aas (2016) for a detailed review). Particularly, Low et al. (2013) reveal the superiority of the minimum expected shortfall portfolio when it was optimized with a dependence function estimated using the canonical vine copula (a special case of the regular vine).

Dissmann et al. (2013) provide an automated algorithm for sequential R-Vine model selection including tree structure, bivariate copula type, and parameter estimation. They also showed that their model provided additional explanatory power over existing dependence models when applied to financial data. In this study, we use the Dissmann et al. (2013) model selection methods, in which each pair-copula can be of a different type chosen from the specified set of 49 bivariate copulas (Algorithm 3.1 in Dissmann et al. (2013)). The algorithm selects the structure in such a way that the highest dependencies are modeled in the first trees. This approach minimizes the influence of rounding errors in later trees, which pairs with strong pairwise dependence are most prone to, e.g., when assessing the joint tail behavior of two variables.

We obtain the real-world R-Vine dependence function using the underlying ETFs' historical return time series. To sample from the joint option-implied distribution, we first sample from the estimated copula model, $\boldsymbol{u} = [u_1, \ldots, u_N]$, where $u \in [0,1]^N$; we then apply inverse marginal cdf estimates to $\boldsymbol{u}$ (inverse sampling). If the cdf is not invertible, as in the case of the Heston model distribution, the joint distribution sample can be obtained using rank correlation.

Appendix E. Historical probability estimate

To benchmark our findings, we also estimate historical marginal probabilities. Suppose that t is the number of trading days and there is a history of n observed daily changes in price logarithms, $I_n = r_t, 0 \leq t \leq n-1$, with $r_{t+1} = \log(S_{t+1}) - \log(S_t)$. The prices, S_{t+1}, are adjusted for ETF splits and dividends. The history, I_n, is used to find the density of the price, $S_{n+m} = S_H$, after another m days of trading, where H is the forecast horizon.

Following the methodology in Rosenberg and Engle (2002) and Taylor (2005), we suppose that an ARCH model for prices is estimated from the history, I_n, and that this model is also applicable in the future. The general historical price process is given by:

$$r_{t+1} = \mu_{t+1} + \varepsilon_{t+1} = \mu_t + \sqrt{V_{t+1}} z_{t+1}, \tag{E.1}$$

where μ_{t+1} is the conditional mean and V_{t+1} is the conditional variance of returns; these conditional time-varying moment functions are determined by the information, I_t, and a parameter vector, θ. z_{t+1} are time-invariant standardised residuals, distributed according to some distribution, $D(0,1)$, with zero mean and unit variance.

Although a constant expected return is not usually compatible with time-varying risk aversion, the effect over a short period (e.g. up to one month) has a negligible impact on probability estimates. Johannes et al. (2002) argue that strategies based on time-varying expected returns perform rather poorly, due to estimation risk, while volatility timing strategies do not suffer from this problem. Therefore, we approximate Eq. (24) by replacing the time-varying mean with constant μ.

On the other hand, multiple studies reveal that variance timing yields a significant economic value for short horizons (see Johannes et al. (2002), Fleming et al. (2003), Carroll et al. (2017)), who find that investors benefit from predicting the dynamics of variance when forming an optimal portfolio). For the conditional variance, we use the asymmetric conditional heteroskedasticity (Glosten et al., 1993) GJR-GARCH(1,1,1) model, which was advocated by Christoffersen and Jacobs (2004b) as a better forecasting model than more richly parameterised models. The choice of the GJR-GARCH model is also determined by the statistical properties of the historical data described in Table 2; in particular, by the presence of ARCH and leverage effects.

The GARCH model is known to weakly converge to a bivariate stochastic model with mean-reverting volatility dynamics (Duan, 1997; Escobar-Anel et al., 2021), which make it a discrete analogue to the Heston model. The leverage effect in the GJR model accounts for the asymmetry of the underlying process. The distribution of the standardised residuals is assumed to be the standardised t-distribution with ν degrees of freedom since the data is heavy-tailed. The GJR-GARCH(1,1,1) model with parameter vector, $\theta = \{\omega, \alpha, \alpha-, \beta, \nu\}$, has the following specification:

$$V_{t+1} = \omega + \alpha\varepsilon_t^2 + \alpha^- \max(0, -\varepsilon_t)^2 + \beta V_t, \tag{E.2}$$

where $\varepsilon_t = \sqrt{V_t} z_t$ and θ is estimated using the maximum likelihood method from history, I_t.

To forecast returns, we first obtain the values of the standardised residuals, $z_{t+1}, \ldots, z_H$, by independent drawings from an estimated Student's t-distribution with ν degrees of freedom. We obtain a 22-period (one month ahead) return by drawing the first return innovation, (z_{t+1}), updating the conditional variance, (V_{t+2}), then drawing the second return innovation, (z_{t+2}), updating the conditional variance, (V_{t+3}), and continuing through to the last innovation. The one-period simulated return, r_H, is equal to $\exp\left[\sum_{i=1,\ldots,22} (\mu + \varepsilon_{t+i})\right]$. We replicate this procedure 10 000 times to obtain the sample input for the SD optimization.

References

Aas, K. (2016). Pair-copula constructions for financial applications: A review. *Econometrics*, 4(4):43.

Ang, A. and Chen, J. (2002). Asymmetric correlations of equity portfolios. *Journal of Financial Economics*, 63(3):443–494.

Atiya, A. and Wall, S. (2009). An analytic approximation of the likelihood function for the heston model volatility estimation. *Quantitative Finance*, 9(3):289–296.

Bawa, V. (1975). Optimal rules for ordering uncertain prospects. *Journal of Financial Economics*, 2(1):95–121.

Carroll, R., Conlon, T., Cotter, J., and Salvador, E. (2017). Asset allocation with correlation: A composite trade-off. *European Journal of Operational Research*, 262(3):1164–1180.

Christoffersen, P. and Jacobs, K. (2004a). The importance of the loss function in option valuation. *Journal of Financial Economics*, 72(2):291–318.

Christoffersen, P. and Jacobs, K. (2004b). Which garch model for option valuation? *Management Science*, 50(9):1204–1221.

Christoffersen, P., Jacobs, K., and Mimouni, K. (2010). Volatility dynamics for the s&p500: evidence from realized volatility, daily returns, and option prices. *The Review of Financial Studies*, 23(8):3141–3189.

Dissmann, J., Brechmann, E., Czado, C., and Kurowicka, D. (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, 59:52–69.

Duan, J. (1997). Augmented garch(p, q) process and its diffusion limit. *Journal of Econometrics*, 79(1):97–127.

Escobar-Anel, M., Rastegari, J., and Stentoft, L. (2021). Option pricing with conditional garch models. *European Journal of Operational Research*, 289(1):350–363.

Fábián, C. I., Mitra, G., and Roman, D. (2011). Processing second-order stochastic dominance models using cutting-plane representations. *Mathematical Programming*, 130(1):33–57.

Fackler, P. and King, R. (1990). The calibration of option-based probability in agricultural commodity markets. *American Journal of Agricultural Economics*, 72:73–83.

Fan, R., Taylor, S., and Sandri, M. (2018). Density forecast comparisons for stock prices, obtained from high-frequency returns and daily option prices. *Journal of Futures Markets*, 38(1):83–103.

Fleming, J., Kirby, C., and Ostdiek, B. (2003). The economic value of volatility timing using "realized" volatility. *Journal of Financial Economics*, 67(3):473–509.

Glosten, L., Jagannathan, R., and Runkle, D. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801.

Hadar, J. and Russell, W. (1969). Rules for ordering uncertain prospects. *The American Economic Review*, 59(1):25–34.

Hanoch, G. and Levy, H. (1969). The efficiency analysis of choices involving risk. *Review of Economic Studies*, 36:335–346.

Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2):327–343.

Johannes, M., Polson, N., and Stroud, J. (2002). Sequential optimal portfolio performance: Market and volatility timing. Available at SSRN 304976.

Kahl, C. and Jäckel, P. (2006). Fast strong approximation monte carlo schemes for stochastic volatility models. *Quantitative Finance*, 6(6):513–536.

Liu, X., Shackleton, M., Taylor, S., and Xu, X. (2007). Closed-form transformations from risk-neutral to real-world distributions. *Journal of Banking & Finance*, 31(5):1501–1520.

Low, R., Alcock, J., Faff, R., and Brailsford, T. (2013). Canonical vine copulas in the context of modern portfolio management: Are they worth it? *Journal of Banking & Finance*, 37(8):3085–3099.

Patton, A. (2004). On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *Journal of Financial Econometrics*, 2(1):130–168.

Rockafellar, R. and Uryasev, S. (2000). Optimization of conditional value-at-risk. *Journal of Risk*, 2:493–517.

Rosenberg, J. and Engle, R. (2002). Empirical pricing kernels. *Journal of Financial Economics*, 64(3):341–372.

Rothschild, M. and Stiglitz, J. (1970). Increasing risk: I. a definition. *Journal of Economic Theory*, 2(3):225–243.

Shackleton, M., Taylor, S., and Yu, P. (2010). A multi-horizon comparison of density forecasts for the S&P 500 using index returns and option prices. *Journal of Banking & Finance*, 34(11):2678–2693.

Taylor, S. (2005). Density prediction for asset prices. In: Asset price dynamics, volatility, and prediction. s.l.:Princeton university press,.