

ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS
Master Thesis Quantitative Finance

Long-term performance analysis of autocallable structured products

Tobias van Steenis (571145)



Supervisor:	Prof. dr. Dick van Dijk
Second assessor:	Dr. Gustavo Freire
Date final version:	30th April 2025

The content of this thesis is the sole responsibility of the author and does not reflect the view of the supervisor, second assessor, Erasmus School of Economics or Erasmus University.

Abstract

Autocallable structured products have grown increasingly popular over the last two decades. However, due to their complexity and lack of long-term historical data, it remains difficult to assess their risk-return characteristics over extended horizons. This research aims to address that gap by providing investors with a clearer understanding of both the potential rewards and the risks involved in a portfolio of autocallables. To overcome data limitations, we develop a novel methodology that combines coupon rate estimation using observable market variables with Monte Carlo pricing under a multi-asset Heston model to construct a single return series for the autocallable portfolio. This portfolio is then compared to an equally weighted portfolio of the underlying indices. The results show that the autocallable strategy consistently outperforms the benchmark on a risk-adjusted basis, primarily by capturing the volatility risk premium. However, this potential for enhanced performance comes at the cost of higher downside risk during extreme market events. These findings highlight a critical trade-off for investors considering an allocation to such products.

Contents

1	Introduction	1
2	Autocallable multi-barrier reverse convertibles	3
2.1	Example	5
3	Data	6
3.1	Implied volatility surface	8
3.2	Method for missing data	9
4	Methodology	12
4.1	Estimating historic coupon rates	13
4.2	Pricing	17
4.2.1	Monte Carlo	18
4.2.2	The single-asset Heston model	19
4.2.3	Calibration of the Heston model	19
4.2.4	Parsimonious multi-asset Heston model	20
4.2.5	Calibration of the multi-asset Heston model	22
4.2.6	Simulation algorithm	23
4.2.7	Variance reduction	24
4.3	Constructing the autocallable strategy	25
4.3.1	Product specification	26
4.3.2	Aggregation into a portfolio	26
4.4	Benchmark	27
4.5	Performance evaluation	27
5	Results	28
5.1	Performance comparison	29
5.2	Autocallable performance: implied vs. realised volatility	33
5.3	Expected shortfall comparison	35
5.4	Coupon rates	37
6	Conclusion	39
7	Appendix	41
A	Tables and Figures	41

1 Introduction

Autocallable multi-barrier reverse convertibles (from now on referred to as 'autocallable' or AMBRC) have emerged as a prominent category of structured financial products. They offer investors enhanced yield potential compared to that of traditional bonds, in exchange for certain conditional risks tied to the performance of underlying assets. Autocallables are composed of a fixed-income part, combined with several exotic options linked to the underlying instruments. They are especially attractive products for investors who seek predictable returns which are higher than bonds, but are prepared to accept losses if the underlying instruments exhibit extreme negative returns. The autocall feature introduces the possibility of early termination of the contract, in exchange for an increase in the coupon rate.

In this paper, we investigate the following question: *What are the risk-return characteristics of a long-term investment in autocallables, and how do they compare to those of a traditional equity portfolio?* To answer this, we construct synthetic return data for autocallable products using a simulation-based approach. This process consists of two main steps. First, we extrapolate historical coupon rates based on observable characteristics of the underlying instruments—specifically the risk-free rate, implied volatility, and correlations between the underlying instruments. This allows us to estimate plausible coupon rates for autocallables at any point in time throughout the analysis period. Second, we evaluate the price evolution of each product using Monte Carlo simulation under a multi-asset Heston model. These two steps allow us to create a portfolio of autocallable notes on a rolling basis, where investments are staggered and reinvested upon maturity, generating a single return series for the strategy. We then compare this strategy to a benchmark: an equally weighted portfolio in the underlying instruments of the autocallables.

Since their introduction in 2003, autocallables have steadily grown in popularity ([Deng et al., 2011](#)), becoming one of the most widely traded structured products in global financial markets. They are predominantly traded over the counter (OTC) but have also gained traction on regulated exchanges in markets such as Germany, Switzerland, and South Korea. According to [Salon \(2019\)](#), autocallables have constituted the most traded class of structured products over the past two decades, reaching an average annual traded volume of 100 billion euros. Their widespread use highlights the need for a better empirical understanding of their investment performance and associated risks.

Despite their popularity, academic research on autocallables remains limited. This is in part due to the structural complexity of these products. Unlike traditional financial instruments, the payoff structure of an autocallable is highly nonlinear, with discrete jumps and path-dependent features. The presence of multiple barriers, early redemption possibilities, and contingent coupon payments makes them infeasible to model analytically,

leaving Monte Carlo simulation as the primary method for valuation. Moreover, the uncertain time to maturity complicates any straightforward comparison to traditional investments.

Beyond modelling challenges, the lack of empirical studies is also driven by data limitations. Unlike stocks or bonds, which have continuous return histories, autocallables are issued as distinct contracts with varying terms, including different issuance dates, maturities, and payoff structures. This makes it difficult to construct a continuous time series of returns over a sufficiently long period, preventing straightforward performance analysis. Consequently, investors have limited empirical guidance on what to expect from long-term exposure to autocallables. This applies both to understanding return dynamics over time and to accurately assessing the associated risks, many of which are not immediately apparent due to the structural complexity of the products.

In the absence of empirical performance research, prior studies have largely focused on the pricing of structured products at issuance. Research on products related to autocallables—specifically (multi-barrier) reverse convertibles—frequently reports overpricing relative to fair value (Wilkins et al., 2003; Szymanowska et al., 2009; Wallmeier & Diethelm, 2009), contributing to a broadly negative academic view of structured products. However, other studies highlight that implied volatility, which is used in option pricing, is systematically higher than realised volatility (Ilmanen, 2012; Israelov et al., 2017). As a result, structured products may appear overpriced even if they ultimately deliver favourable realised outcomes. The existence of this volatility risk premium is often linked to behavioural finance insights, such as loss aversion and the overweighting of small-probability events, which drive investor demand for protective features embedded in these products (Kahneman & Tversky, 1979). Together, these findings suggest that autocallables may offer more attractive risk–return characteristics than pricing-based assessments imply, a possibility that underscores the relevance of our empirical investigation.

The results of this research are consistent with this hypothesis: a rolling portfolio of autocallable structured products consistently outperforms a traditional equity benchmark on a risk-adjusted basis. The strategy benefits from capturing the volatility risk premium, as market-implied probabilities of knock-in events are consistently overstated relative to realised outcomes. However, this potential for enhanced performance comes with higher downside risk during extreme market events, as the expected shortfall of the autocallable portfolio is, on average, nearly 10% higher than the benchmark. This difference is primarily attributed to the “worst-of” feature embedded in these products. These findings highlight a critical trade-off for investors considering an allocation to autocallables.

We contribute to the literature by investigating the long-term performance of autocallables compared to more traditional investments. This offers practical insights for investors on whether these products are a worthwhile investment, as well as how they could be integrated into a broader strategy. Additionally, by investigating this topic we develop a more

general step-by-step framework through which one can examine the investment potential of autocallable multi-barrier reverse convertibles. This is also applicable to several closely related types of structured products, for example: regular (multi) barrier reverse convertibles. By providing this empirical foundation and methodological approach, we aim to facilitate further academic research into structured products. One natural extension would be to explore portfolio optimisation within a structured product framework, where an investor could make dynamic allocation decisions across different product architectures, barrier levels, or underlying indices to maximize expected utility or other objective functions.

The remainder of this paper is structured as follows. Section 2 provides an in-depth discussion of autocallables and outlines their key components. Section 3 presents the data used in this study, followed by Section 4, which describes the methodological approach. Section 5 then presents the main findings, and finally, Section 6 concludes the paper.

2 Autocallable multi-barrier reverse convertibles

In Section 1 we have briefly explained what an autocallable multi-barrier reverse convertible is. This section provides a more in-depth overview of the product. As mentioned, autocallables offer periodic coupon payments that are significantly higher than those of traditional bonds. Moreover, the underlying structure of an AMBRC involves setting multiple barrier levels and an autocall feature, both of which are based on the performance of a specified underlying asset, such as a stock, index, or a basket of securities. We explain how the product is constructed using a combination of fixed-income instruments and exotic derivatives, and how these components influence its payoff structure.

Firstly, an AMBRC consists of a simple zero-coupon bond. Secondly, it contains a written knock-in put option. This is an option that activates only if the underlying asset's price falls below a predetermined barrier level during the option's life, the so-called knock-in barrier. Before this barrier is reached, the option does not exist, and if the barrier is never touched, the option expires worthless. However, if the barrier is breached, the knock-in put becomes a standard put option, allowing the holder to sell the asset at a specified strike price. The strike price is usually set at the current price of the underlying, and the knock-in barrier at a certain percentage of the strike, usually in the range of 50-60%. This component is the only source of downside risk of an autocallable, which means that the product is partially capital protected. If the knock-in barrier is breached during the product's lifetime, the payoff of the product is equal to the value of the underlying at maturity, likely resulting in large losses. In any other case the product is principally protected with the possibility of additional coupon payments.

These coupon payments compensate the investor for the risk taken by selling the knock-in put option to the issuer. They are determined by another component of the

AMBRC, namely digital call options. This type of option provides a fixed payout if the underlying asset's price is above a specified strike price at the option's expiration. Embedded in the product is a series of digital options, each with a distinct maturity corresponding to an observation date. The strike for these options, known as the coupon barrier, is also commonly set at a certain percentage of the current price of the underlying, but it is always higher than the knock-in barrier. This barrier usually falls in the range of 70-95%. Importantly, the underlyings of the digital call options are the same as those of the knock-in put, ensuring that coupon payments are contingent on the performance of the same instruments that determine downside risk. The coupon barrier is observed periodically, and at each observation date it is determined whether the coupon will be paid. It should be noted that the occurrence of a knock-in event does not affect the coupon payment; the coupon barrier is the sole determinant of whether a coupon is paid. Moreover, if the coupon is withheld because the underlying asset falls below the coupon barrier, and later, on another valuation date, the underlying asset rises above the coupon barrier, the coupon payment is made as usual, along with any previously unpaid coupons. This feature is known as the "memory" effect of the product.

The zero-coupon bond, the written knock-in put, and the digital calls, together constitute a (regular) multi-barrier reverse convertible. All that remains is the autocall feature, which means that the product can be automatically redeemed before its maturity date if the underlying asset's price reaches or exceeds a predetermined level, known as the autocall barrier, on specified observation dates. This is achieved by means of a written Bermudan call option on the entire multi-barrier reverse convertible. This way the issuer of the product can buy back the AMBRC if the underlying instruments are above the autocall barrier on an observation date. The autocall barrier is often around the current level of the underlying instrument.

As for the underlying assets of the AMBRC, they are often equity products, ranging from single stocks to popular equity indices across different markets. In most cases, these products are based on a basket of underlying assets instead of a singular asset. The payoff of the structure is then based on the worst-performing asset of the basket. This is commonly referred to as the "worst-of" feature. In this paper we will solely deal with these worst-of autocallables, as they are the most common. However, the methods are easily generalizable to the case where the autocallable is based on just one asset.

The payoff structure of AMBRCs is particularly appealing to investors seeking higher income through elevated coupon rates, but it also requires them to carefully consider the associated risks caused by the possibility of a knock-in. The principal protection is conditional upon the underlying asset's price movements, making it crucial for investors to understand the potential scenarios in which their initial investment may be jeopardised.

2.1 Example

To further illustrate the payoffs of an autocallable, we present a series of examples that demonstrate key scenarios. These examples pertain to a hypothetical autocallable product with the following characteristics: the underlying basket comprises three indices, with the product's payoff based on the worst-performing index. The nominal amount of the autocallable is 1,000,000 EUR with a maximum maturity of one year. The product pays a coupon rate of 10% annually. The product features a knock-in barrier set at 60%, a coupon barrier at 70%, and a constant autocall barrier at 90%. The coupon and autocall barriers are observed only on four specific observation dates. The knock-in barrier is American, meaning that if the underlying breaches the barrier at any point during the product's lifetime, the written put option is activated. Additionally, the product incorporates a memory coupon feature.

The first example is shown in Figure 1. In this case, the worst-performing index remains below the autocall barrier throughout all periods, meaning the product continues until the maximum maturity of one year. At the end of the first two periods, the worst-performing index is above the coupon barrier, resulting in a coupon payment of 2.5% at both observation dates, amounting to 25,000 EUR at each date. However, during the third period, Index A breaches the knock-in barrier, meaning the repayment at maturity will be tied to the price of the worst-performing index at that time. With Index A priced at 47.6 at maturity, the resulting payment is 476,000 EUR. It is important to note that coupons for periods 3 and 4 could still have been paid, despite the knock-in barrier breach, if the worst-performing index had risen above the coupon barrier on the respective observation dates. However, this was not the case in this example.

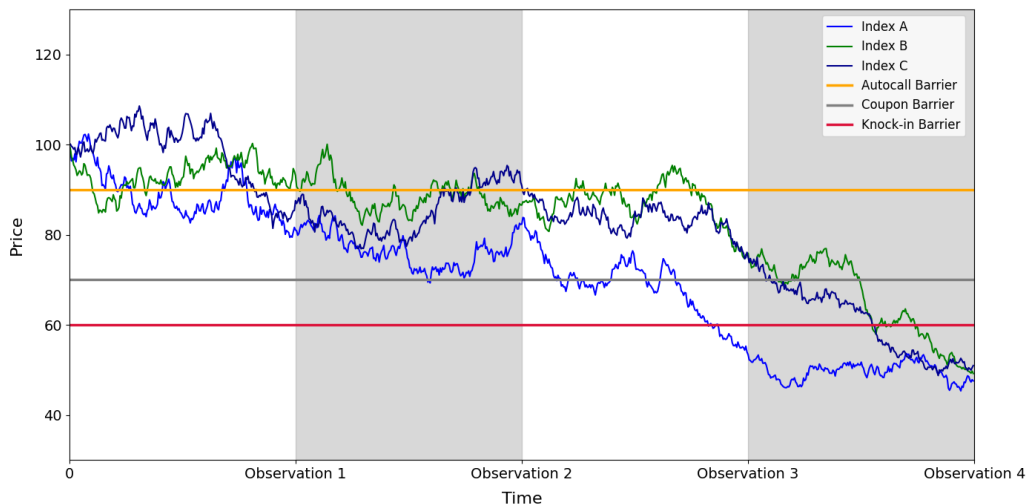


Figure 1: Price paths of the underlying indices in example 1

The second example is shown in Figure 2. On the first observation date, the worst-performing index, Index A, falls below the autocall barrier, meaning the product remains

active. Additionally, it breaches the coupon barrier, resulting in no coupon being paid at the end of the first period. In the second period, the worst-performing index, now Index C, remains below the autocall barrier but rises above the coupon barrier. As a result, the coupon for the second period is paid at 2.5% of the principal. Due to the memory coupon feature, all previously unpaid coupons are also paid, meaning the investor retroactively receives the coupon from the first period at the end of the second period. In the third period, all indices are above the autocall barrier, triggering the early termination of the autocallable. Consequently, the principal amount of 1,000,000 EUR is repaid to the investor, along with the coupon payment for the third period.

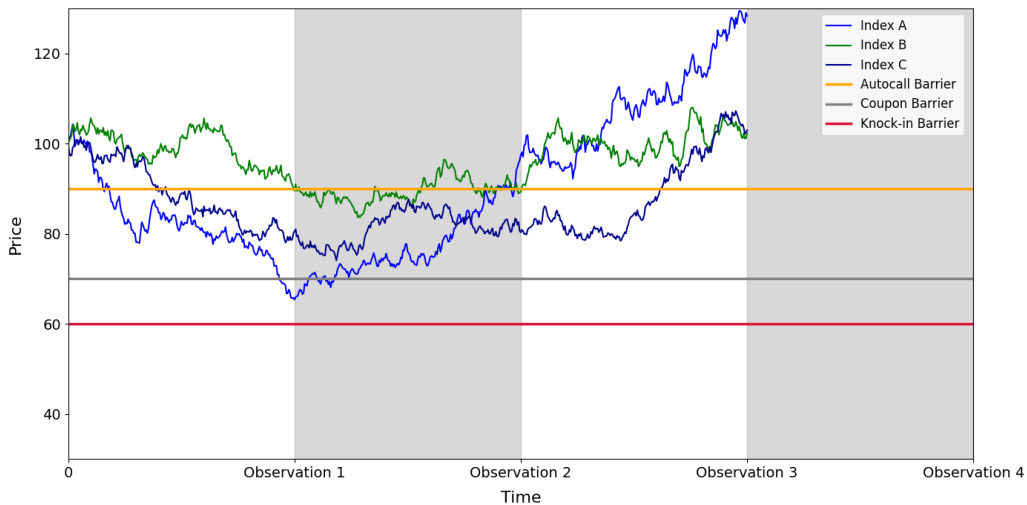


Figure 2: Price paths of the underlying indices in example 2

3 Data

In this section, we discuss the origin and characteristics of our dataset. As discussed previously, research on the performance of portfolios containing structured products is sparse primarily because the necessary data is particularly difficult to obtain. There are no public databases which contain historical returns series of autocallable products over a long period of time. The dataset for this research is provided by the Market Stability Fund (MSF), which constructs the portfolio of their fund around OTC autocallable products. The dataset contains all of the 342 products that the firm has traded during its existence (from January 2018 until April 2024). The information on these products includes the trade and maturity date, the annual coupon rate, the payment barriers (knock-in barrier, coupon barriers and autocall barriers) of each autocallable as well as the underlying basket of the product.

Barring a few exceptions, the autocallable products in the dataset utilise five underlying instruments, with the final payment at maturity determined by the worst-performing asset in the basket. We delete the four observations where the basket contains less than

five underlyings. The underlying instruments are exclusively well-known stock indices from various markets, ensuring broad cross-sectional diversification and avoiding the idiosyncratic risk associated with individual stocks. A total of 21 different indices appear across these baskets. However, 11 of these indices are included in the baskets at most 11 times. As a result, structures that contain these less frequent indices are removed from the dataset to avoid potential overfitting and ensure more robust modelling. The remaining 10 indices included in the analysis are: Euro Stoxx 50, S&P 500, FTSE 100, FTSE MIB, Nikkei 225, Russell 2000, DAX, CAC 40, Hang Seng Index and MSCI Emerging Markets Index.

Almost all of the autocallables use a knock-in barrier of 60%, meaning that if the worst-performing underlying index falls below 60% of its initial value at the trade date, the product will return the value of the worst-performing asset at maturity instead of the original principal. In addition, the coupon barrier for most autocallables is set at 70%, meaning that periodic coupon payments depend on the worst-performing index staying above 70% of its initial value. All products utilise a "memory" coupon, as explained in Section 2. The autocall barriers differ between products, but are almost always in the 90-100% range of the strike price. Some products utilise a constant autocall barrier across each observation date, while others utilise a step-down methodology. This means that the autocall barrier decreases over time (e.g. with 1% per month), making it easier for the product to be called as the observation periods progress.

The autocallable products have a maximum maturity ranging from 1 to 5 years, with the majority maturing within the 1-2 year range. The performance analysis in this research will focus on autocallables with a maximum maturity of 1 year, as this structure allows for the most accurate and consistent pricing. To ensure the integrity of the analysis, all products with a maximum maturity of over 2 years are excluded from the dataset. This is because longer-term products often exhibit coupon rates that differ significantly from those of shorter maturities, potentially skewing the results and leading to inaccurate conclusions about performance and pricing. Removing these longer-maturity products ensures a more focused and reliable analysis of autocallables with similar characteristics. After deletion there remain 286 structures in the dataset. Among these products, the annual coupon rates exhibit considerable variation, with a mean of 11.6%, a standard deviation of 2.3%, and values ranging from a minimum of 6.7% to a maximum of 18.4%. This variation reflects differences in market conditions, underlying basket composition, and product design.

To provide a more complete picture of the remaining dataset, Figure 3a shows the number of times each index is included in a product. Additionally, Figure 3b illustrates the evolution of the coupon rate for the autocallables in the dataset, with the VIX included as a reference to market conditions during each period. The chart clearly highlights the correlation between market volatility and the corresponding coupon rates. The figure

displays the 5-observation rolling window for both the coupon rate and the VIX, applied to reduce noise and better highlight the trends over time.

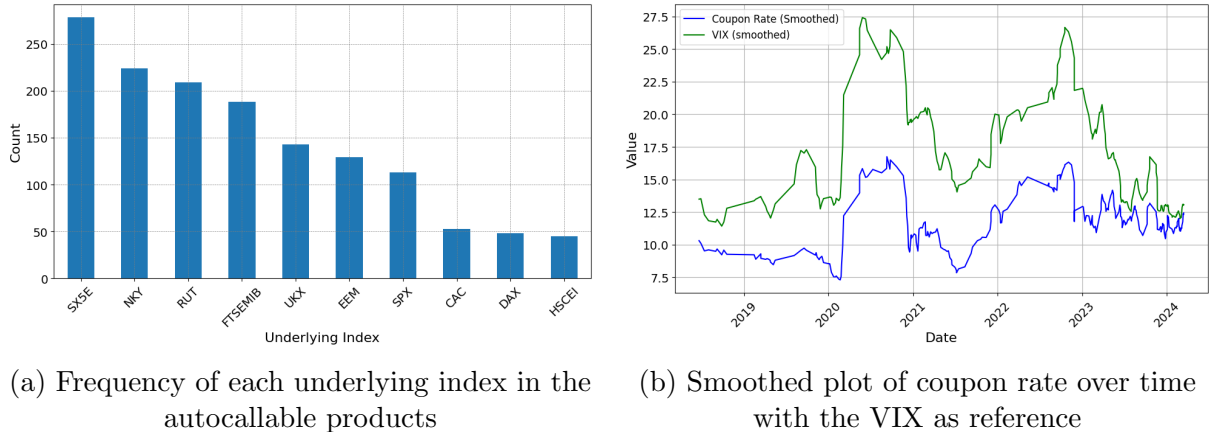


Figure 3: Summary figures of the autocallable products

In addition to data on the autocallable products, we utilise information on the underlying instruments, including index returns, implied volatilities, corresponding risk-free rates, and dividend yields. For implied volatility in the historic coupon estimation, we rely on 30-day implied volatility indices, such as the VIX for the S&P 500 and the VSTOXX for the Euro Stoxx 50. Risk-free rates are matched to the respective markets: we use the 3-month EURIBOR for European indices and the 3-month TIBOR for the Nikkei 225. For UK indices, we apply the 3-month GBP LIBOR until its discontinuation at the end of 2021, transitioning to SONIA thereafter in line with prevailing market practices. Similarly, for US-based indices, we use the 3-month USD LIBOR until the end of 2021, after which we adopt SOFR as the benchmark risk-free rate. This data is sourced from Bloomberg. The research period spans from April 7, 2006, to April 15, 2024. The starting date is determined by the availability of TIBOR data.

Index returns are analyzed at a daily frequency to account for potential knock-in events, as the autocallables feature American-style knock-in barriers that require daily monitoring. However, the performance of the investment strategies is evaluated on a weekly basis. This approach balances computational feasibility with analytical accuracy, as the pricing algorithm for autocallables is computationally intensive. Using a finer time frame, such as daily intervals, could artificially inflate volatility estimates due to the accumulation of simulation errors inherent in the pricing process. By transitioning to weekly intervals, these errors are significantly reduced while maintaining sufficient observations for robust analysis.

3.1 Implied volatility surface

The pricing of autocallable products requires modelling the entire implied volatility surface (IVS). The IVS represents implied volatility across different strike prices and maturities for

a given underlying asset. Typically, the IVS exhibits a “smile”, where volatility increases for options far from the current market price. It also varies across maturities, with the term structure often positive in low-volatility regimes, indicating expectations of rising long-term volatility. Conversely, during high-volatility regimes, the volatility is expected to mean-revert in the long-term, resulting in a negative term structure.

Modelling the full IVS is essential because autocallables are highly path-dependent, incorporating features such as American-style knock-in barriers and the memory effect on coupon payments. These characteristics make the payoff sensitive to the entire trajectory of the underlying instruments, rather than just their values at predetermined observation dates. In addition, our portfolio includes products with varying barrier levels and observation schedules, further increasing the need for accurate simulation across a broad range of potential outcomes. Consequently, a complete and smooth IVS enables the pricing model to generate market-consistent valuations across diverse paths and product structures.

To obtain this data, we use the Ivy DB US OptionMetrics database from Wharton Research Data Services, which provides IVS data for US-based indices, including the S&P 500, Russell 2000, Nasdaq 100, and Dow Jones Industrial Average. However, IVS data for other indices required for our analysis is unavailable. In the following section, we describe the method used to address this gap.

The dataset includes maturities ranging from 10 days to 2 years. We exclude the 10-day maturity, as stochastic volatility models often struggle to fit the short end of the IVS, potentially skewing calibration. Given that the first autocall observation for the products occurs after one month, accurately fitting the 30-day implied volatility is sufficient for precise pricing. Maturities longer than one year are also excluded, as they exceed the maximum maturity of the autocallables modeled in this study. The remaining maturities are 30 days, 60 days, 91 days, 122 days, 152 days, 182 days, 273 days, and 365 days.

For strike prices, the dataset utilises a delta-based surface. Following [Farkas et al. \(2024\)](#), we focus on out-of-the-money (OTM) options due to the illiquidity of in-the-money (ITM) options caused by infrequent trading. Specifically, we include deltas from -15 to -50 for puts and from 45 to 15 for calls, in increments of 5. The IVS dataset ends on 25 August 2023, which consequently serves as the final observation date used in this analysis.

3.2 Method for missing data

As we only have access to the IVS for US based indices, we need a method to leverage this data to obtain reliable estimates for the IVS of other equity indices on which the autocallable products are based. The required indices include the Eurostoxx 50 (SX5E), the FTSE 100 and the Nikkei 225.

Some data on the volatility of these required indices is available, namely their respect-

ive volatility benchmarks, such as VSTOXX for the SX5E and VFTSE for the FTSE 100. These benchmarks are equivalent to the VIX for the S&P 500 in that they measure the market’s expectations of future volatility over a 30-day horizon using a weighted average of the 30-day implied volatilities across different strike prices. Each index’s volatility benchmark reflects regional economic conditions, market sentiment, and the liquidity of the underlying options market.

A practical approach to approximate the missing surfaces is to scale an available surface (such as that of the S&P 500) by the ratio of volatility benchmarks. For instance, to estimate the IVS of the SX5E, one would scale the S&P 500 IVS by the ratio $\frac{\text{VSTOXX}}{\text{VIX}}$. This method leverages the information embedded in benchmark volatility indices, which reflect market expectations of volatility for each region or asset class, thereby providing a relative measure of expected volatility between the two markets. The effectiveness of this approach relies on the assumption that most differences in IVS across indices stem from differences in the absolute level of implied volatility (as captured by the benchmark indices), while the overall shape—such as the smile and term structure—remains broadly consistent across indices. However, upon deeper investigation, this assumption was found to be only partially valid:

1. Smile: [Mixon \(2011\)](#) states that normalized skew is the most descriptive measure of the shape of the smile. Normalized skew is calculated as the difference between the 25D put and 25D call implied volatilities, divided by the ATM implied volatility. The reasoning behind this statement is that this measure is mostly independent of the absolute level of the IVS. This also holds true in our analysis as illustrated in [Figure 4a](#). There, the normalized skew is regressed on the level of the IVS for the Russell2000. One can observe that the level exhibits minimal explanatory power on the normalized skew, which is confirmed by the R^2 of the regression, which is only 0.03. The simple scaling method as described above, ensures that the normalized skew is preserved, as the put-call differences in implied volatility relative to the ATM IV are unchanged. This makes the scaling effective for modelling the smile of the estimated IVS.
2. Term Structure: Unlike normalized skew, the term structure of implied volatility—characterized by differences in IV across maturities—exhibits significant dependence on the level of implied volatility, as illustrated in [Figure 4b](#). This figure shows the difference in average IV (i.e. averaged across deltas) between the 365-day and 30-day tenors. A clear pattern emerges: as the level of the IVS increases, the differences in volatility between the two maturities become more negative. At lower overall volatility levels, longer-term volatility tends to exceed short-term volatility, resulting in an upward-sloping term structure. Conversely, at higher volatility levels, the term structure slopes downward, reflecting the expectation that elevated volat-

ility regimes are unlikely to persist over extended periods. These findings highlight the mean-reverting nature of volatility. Consequently, the term structure cannot be considered independent of the level of implied volatility, making the simple scaling approach insufficient to accurately capture these variations. The findings for the relationship of the level with both the smile and term structure are very similar across the different indices. The complete results can be found in Appendix A.

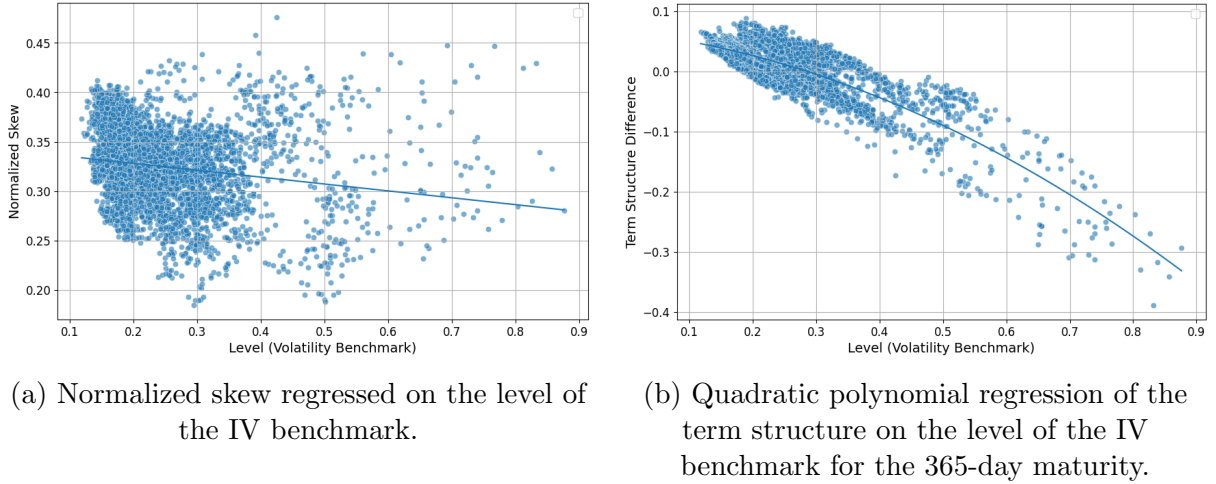


Figure 4: Test for the independence between the shape and level of the Russell 2000 implied volatility surface

To account for the dependence between the term structure and the level of the IVS, we propose a method that builds on the relative scaling of the IVS by the volatility benchmarks of indices i and j . This scaling ensures the baseline adjustment reflects the relative levels of implied volatility. In addition to this, the method incorporates a term structure adjustment based on a quadratic polynomial regression, as shown in Figure 4b, to capture systematic relationships between the level and term structure. A quadratic regression is chosen over a linear specification, because the latter was found to structurally underestimate the steepness of the term structure in extreme volatility regimes. Since these high-volatility observations have a large impact on the performance of the autocallables, it is crucial that they are modeled as accurately as possible. Finally, the observed deviations from the regression for index i are transferred to index j . As global indices are highly correlated, these deviations are expected to exhibit similar patterns between indices, providing a flexible way to reflect idiosyncratic term structure features. For instance, during global events such as the banking crisis or COVID, we can reasonably expect term structures across indices to behave similarly. This combined approach ensures that the scaled IVS for index j accurately captures both systematic term structure relationships and the unique characteristics of index i at any time. Importantly, the linear scaling process inherently preserves the normalized skew, maintaining consistency in the moneyness structure.

The target level of the implied volatility for a specific maturity T_k is calculated as:

$$IV_{j,\text{target}}(T_k) = IV_i(30) \cdot \frac{V_j}{V_i} + f_{i,T_k}(V_j) + (IV_i(T_k) - IV_i(30) - f_{i,T_k}(V_i)), \quad (1)$$

where:

- $IV_i(30)$ is the average 30-day implied volatility of the surface for index i , representing its short-term volatility level.
- V_i and V_j are the volatility benchmarks (e.g., VIX, VSTOXX) for indices i and j , respectively. $IV_i(T)$ is calculated as the average implied volatility across deltas for a given maturity T .
- $f_{i,T_k}(\cdot)$ is the regression model that estimates the term structure difference $IV_i(T_k) - IV_i(30)$ for index i as a function of its volatility benchmark V .
- The final term $IV_i(T_k) - IV_i(30) - f_{i,T_k}(V_i)$, represents the residual of the term structure regression for index i . This ensures that the idiosyncratic features of index i are transferred to the predicted IVS for index j .

The IVS is then scaled per maturity. Specifically, The implied volatilities corresponding to maturity T_k are scaled proportionally to the ratio of the target level to the original average level for that maturity:

$$\text{Scaling Factor}_{T_k} = \frac{IV_{j,\text{target}}(T_k)}{IV_i(T_k)}. \quad (2)$$

Finally, our dataset includes IVS data for four indices: the S&P 500, Russell 2000, Dow Jones, and Nasdaq 100. To enhance the robustness of our method for estimating unknown volatility surfaces and reduce sensitivity to potential outliers in any single index, we leverage the IVS data of all four indices. Specifically, we estimate the missing surfaces individually using each of these indices as a reference and then take the average of the results. This averaging process ensures a more stable and reliable estimate by mitigating the influence of idiosyncratic deviations in any one index.

4 Methodology

As discussed in Section 1, the most challenging part of this research lies in constructing the return series for the autocallable products. The available dataset only includes products issued between 2018 and 2024, a relatively short period that limits the ability to draw robust conclusions about the long-term performance of autocallables relative to traditional equity investments. To conduct a more thorough analysis, it is therefore necessary to extend the timeframe over which the strategy can be evaluated. This involves a three-step

process. First, we estimate plausible coupon rates of the autocallable for each week over the lengthened analysis period. The second step is to calculate the prices of these products over their lifetime. The final step is to put these products together into a portfolio, referred to as the autocallable strategy. We then introduce the benchmark portfolio used for comparison and conclude the section by describing the methods employed to evaluate the performance of both strategies.

4.1 Estimating historic coupon rates

As we know from classic portfolio theory, the reward of an investment should be proportional to its risk. Thus any model for the coupon rate of an autocallable should be determined by its risk factors. In the absence of prior research in modelling the coupon rates of structured products, we follow a standard approach in the field of risk analysis, where it is common to utilise a linear factor model to model the risk ([Anderson-Cook, 2006](#)). The frequent use of linear factor models can be attributed to their simplicity and interpretability. Another approach to modelling the coupon rates would be to calibrate them such that the product's fair value at issuance equals its nominal value. This would ensure that no capital transfer is required at inception, which aligns with what is typically observed in the market. However, this approach has notable drawbacks. First, it makes the resulting coupon rates highly sensitive to the choice of pricing model, which introduces model risk. Second, as observed in the literature and demonstrated in this research (see [Section 5](#)), autocallables most frequently exhibit a fair value below par at issuance. Setting the coupon to enforce par valuation would therefore lead to inflated coupon rates, creating a bias in favour of the autocallable strategy. For these reasons, we instead choose to estimate coupon rates based on observed market data and actually traded products, ensuring a more realistic and representative input for the analysis.

The risks of an autocallable are entirely dependent on its underlying instruments, specifically when the worst performing instrument drops below the knock-in barrier, as in this case the product returns the value of this instrument, resulting in a loss. Even though there is little data on structured products, there is no shortage of data on the underlying stock indices. A natural indicator of the risk of these stock indices is their implied volatility, as a higher implied volatility results in a higher likelihood to breach the knock-in barrier. However, the implied volatilities do not paint the full picture, as the movement of stock indices exhibit large co-movements, which impacts the probability of a knock-in as well. Large correlations between assets mean that if one of the assets is to breach the knock-in barrier, it is more probable that other assets will mirror this movement, while in all other cases the barrier will not be breached. On the other hand, for stocks with low or zero correlation, their independent movements result in a mix of performances within the basket. This means that the probability of at least one stock

performing poorly enough to breach the knock-in barrier becomes greater. To illustrate this argument, compare the case where the underlying instruments exhibit perfect correlation with the case where they are uncorrelated. In the first case, the probability of the security hitting the knock-in barrier is equivalent to the knock-in probability of the asset with the highest volatility. This is because the assets move in unison, so the security's behavior is driven by the most volatile asset. Conversely, when the assets are uncorrelated, each asset has an independent chance of breaching the barrier. As a result, the combined probability of at least one asset hitting the barrier is higher, leading to an increased overall probability of a knock-in event.

This effect is more difficult to capture in the model, as correlations are not directly observed and thus have to be estimated. There also do not exist straightforward analytical solutions to estimate the effect of these correlations. For these reasons we turn to Monte Carlo methods to incorporate the effects of correlation on the coupon rate of the product. We do this by simulating the price paths of the underlying instruments of each autocallable note, for which we use correlated geometric Brownian motions (GBMs) as described in [Ang \(2015\)](#). Then, for each iteration we can observe if one of the underlying instruments breaches the barrier, meaning that a knock-in event has occurred. Finally we get our estimate of the knock-in probability as the proportion of iterations where a knock-in event has occurred. The empirical correlation matrix for a time series of returns of K observations is defined as $\Sigma_K^{\text{emp}} = (\rho_{i,j:K}^{\text{emp}})_{1 \leq i,j \leq d}$. The elements of Σ_K^{emp} are calculated as

$$\rho_{i,j:K}^{\text{emp}} = \frac{\sum_{k=1}^K (R_i(t_k) - \bar{R}_i)(R_j(t_k) - \bar{R}_j)}{\sqrt{\sum_{k=1}^K (R_i(t_k) - \bar{R}_i)^2} \sqrt{\sum_{k=1}^K (R_j(t_k) - \bar{R}_j)^2}}, \quad (3)$$

where $R_i(t_k)$ is the return of asset i in period t_k and $\bar{R}_i = \frac{1}{K} \sum_{k=1}^K R_i(t_k)$ is the mean return over the observation period. We use $K = 252$, corresponding to the typical number of trading days in a year, which aligns with the maximum maturity of the autocallable products analyzed in this study. This correlation matrix is then used as input for the simulation of the correlated brownian motions of the underlying indices of the autocallable products.

A limitation of this approach is that it does not account for the different time zones of the underlying indices when calculating correlations. For example, the Nikkei 225 and S&P 500 have non-overlapping trading hours, which introduces potential lead-lag effects in the return data. This misalignment tends to reduce estimated correlations, making the indices appear more independent than they are in reality. As a result, the Monte Carlo simulation may overstate the probability of a knock-in event. However, since these probabilities are consistently estimated across all products and are used as relative explanatory variables in the regression model, the effect on the estimated coupon rates is unlikely to result in systematic bias. While the absolute levels of knock-in probability may

be inflated, their explanatory power for coupon setting remains intact. It is worth noting, however, that this limitation may be more consequential in the pricing phase, where the correlations directly affect simulated payoffs. We return to this point in Section 4.2.4 when discussing the Heston-based pricing model.

The final component of the coupon model is the risk-free rate. While not a risk factor in itself, it sets the baseline return for risk-free investments. When interest rates rise, safer alternatives like government bonds become more attractive, and structured products must offer higher expected returns to remain competitive. Additionally, higher rates lower the present value of the zero-coupon bond component embedded in the autocallable, which reduces the cost of funding the principal repayment. This allows the issuer to allocate more value to the coupon and option components, ultimately resulting in higher offered coupon rates.

Following these principal risk factors of autocallable products, we construct a linear regression model that maps these risk factors to the estimated coupon rate of an autocallable at any given point in time. Specifically, the coupon rate for an autocallable i , with trade date t , can be calculated as:

$$Coupon_i = \alpha + \beta r_t + \gamma^T \bar{\sigma}_i + \delta p_i, \quad (4)$$

where $Coupon_i$ is the annual coupon rate. α is a constant, r_t is the risk free rate at time t with coefficient β , γ is the vector of coefficients for the implied volatilities of the 10 different available indices. δ is coefficient for the knock-in probability and p_i is the probability of a knock-in event for autocallable i as estimated by the Monte Carlo simulation. For the risk-free rate, we use the 3-month EURIBOR, as the majority of autocallables in the dataset are denominated in euros. The implied volatility vector $\bar{\sigma}_i$ warrants extra explanation. As detailed in Section 3, each autocallable has a mix of five underlying indices. So in order to estimate its coupon rate, only those five indices should be taken into account. Thus to obtain the implied volatility vector of autocallable i , we multiply the implied volatilities at the trade date t elementwise with a vector of dummy variables which indicate whether the corresponding index is included in the underlying instruments of the product. The result of this transformation is that the implied volatility remains the same if its index is in the underlying instruments, and zero if it is not. Mathematically, this is represented as: $\bar{\sigma}_i = \sigma_t \circ \mathbf{I}_i$. Here, $\sigma_t = (\sigma_{t,1}, \sigma_{t,2}, \dots, \sigma_{t,n})$ is the vector of all implied volatilities, $\mathbf{I}_i = (\mathbb{1}_{\{1 \in A_i\}}, \mathbb{1}_{\{2 \in A_i\}}, \dots, \mathbb{1}_{\{n \in A_i\}})$ is the indicator vector selecting the relevant implied volatilities and A_i is the set of index identifiers (e.g. 1 through n) included in the underlying basket of autocallable i . This construction filters the volatility vector σ_t , retaining only the volatilities of the indices relevant to autocallable i . The resulting vector is then used as an input in the regression model.

A potential concern in this specification is the presence of multicollinearity among the

implied volatilities, as equity index volatilities tend to be strongly correlated, particularly during periods of market stress. However, this issue is mitigated by the structure of the dataset: autocallable products in the sample exhibit a wide variety of underlying index combinations. This cross-sectional variation in basket composition enables the model to better disentangle the individual contributions of each index’s implied volatility. As a result, despite the inherent correlation structure, the regression is still able to identify the relative impact of each index’s risk exposure on the offered coupon rate.

Another point that warrants clarification is the difference in underlying processes used for the regression versus the pricing model. While we use a geometric Brownian motion framework to simulate knock-in probabilities for the regression, the pricing model in Section 4.2 relies on a more sophisticated Heston model with stochastic volatility. This choice is motivated by the role of the knock-in probability in the regression: it serves as a relative measure of risk that helps explain variation in coupon rates across products, rather than needing to represent its exact, real-world value. As long as the same model is applied consistently across the sample, these relative differences remain valid. Moreover, the GBM framework enables us to incorporate real, directly observable implied volatility data for all products. In contrast, the Heston model requires a full implied volatility surface, which in our case must be estimated for eight out of ten indices, using the method described in Section 3.2. Using GBM in this context reduces estimation risk and provides more reliable inputs for the regression, making it the more practical and robust choice. For pricing, however, the underlying process must more accurately reflect true risk-neutral probabilities, as it directly determines the valuation of future payoffs.

In order to test the adequacy of the coupon estimation approach, we can compare the in- and out-of-sample performance metrics of the regression. The in-sample R^2 is 0.68 with a mean absolute error (MAE) of 1.03%. This means that the predicted coupon rate is on average 1.03% removed from the actual coupon rate. While this level of error is not negligible in isolation, it constitutes a substantial improvement over a naive model assuming a constant coupon rate. This is especially evident when considered in light of the characteristics of the dataset, as discussed in Section 3: the average coupon rate is 11.6%, with a standard deviation of 2.3 percentage points and values ranging from 6.7% to 18.2%. The substantial cross-sectional variation implies that any constant-rate approximation would be poorly calibrated across most observations.

Furthermore, it is unlikely that the model’s error can be significantly reduced. For instance, we observe that two autocallables traded on the same date with identical underlying indices can still differ by more than two percentage points. This variation reflects the fact that issuing banks apply different pricing models and assumptions, resulting in differing views on the fair coupon rate. In the final analysis, however, these differences should not have a large effect on the outcome, as the coupons will sometimes be overestimated, and sometimes underestimated, averaging out in the end. The out-of-sample performance

is tested by splitting the sample into an 80-20 train-test split. We estimate the model on the training sample and calculate the performance metrics for the test sample. We show the average performance of a 1000 random train-test splits. The average R^2 drops to 0.63 and the MAE increases to 1.09%. This shows that the model generalizes quite well to unseen data, as the out-of-sample performance decreases only slightly compared to in-sample.

4.2 Pricing

An integral component of this research lies in accurately pricing autocallable products, as reliable valuations are essential for assessing their long-term return and risk characteristics. If pricing is over- or underestimated due to poor modelling, this may lead to unrealistic expectations about the product’s performance, either overstating potential upside or ignoring potential risks.

Various pricing frameworks have been proposed in the literature, each involving trade-offs between computational efficiency, realism, and suitability for complex payoff structures. Analytical pricing methods, which provide closed-form solutions, are computationally efficient and offer clear insight into model behaviour. However, they lack the flexibility to accommodate the features of modern structured products. For example, the approaches used by [Deng et al. \(2011\)](#) and [Guillaume \(2015\)](#) are not designed to handle multi-asset, worst-of structures that include elements such as memory coupons or American-style knock-in barriers. Some Monte Carlo methods, such as those by [Glasserman & Staum \(2001\)](#) and [Alm et al. \(2013\)](#), aim to reduce computational burden, but rely on the Black-Scholes assumption of constant volatility, which limits pricing accuracy, as it does not capture known market dynamics such as the volatility smile.

To address these limitations, we adopt a stochastic volatility (SV) framework, originally introduced by [Heston \(1993\)](#), which more accurately captures the volatility dynamics observed in financial markets. In this model, volatility follows its own stochastic process and is no longer assumed to be constant. This is particularly relevant for autocallables, which derive a substantial portion of their risk from a deeply out-of-the-money knock-in put option. In such cases, constant-volatility models tend to underestimate the risk of tail events, as they cannot capture the sharp increases in volatility typically observed during market stress. In contrast, stochastic volatility models are better equipped to reflect these dynamics, providing a more accurate representation of the risk embedded in these products. To extend this framework to a multi-asset context, we implement the parsimonious version of the multi-asset Heston model proposed by [Dimitroff et al. \(2011\)](#). This formulation captures essential cross-asset dependence while avoiding the instability and computational burden associated with estimating a large number of asset-volatility and volatility-volatility correlation parameters.

Several alternative frameworks have also been explored in the literature. The local volatility (LV) model, introduced by [Dupire \(1994\)](#), is widely used due to its ability to fit the implied volatility smile exactly. However, as it treats volatility as a deterministic function of time and asset price, it fails to capture the stochastic behaviour of volatility over time. As a result, it tends to misprice volatility-of-volatility risk, as noted by [Bergomi \(2015\)](#). More recently, local-stochastic volatility (LSV) models have been proposed as a hybrid solution that combines the strengths of LV and SV approaches ([Farkas et al., 2024](#)). While effective in capturing short-term volatility dynamics, LSV models are computationally intensive and generally limited to single-asset applications, as highlighted by [De Col & Kuppinger \(2017\)](#).

Given these considerations, the stochastic volatility approach strikes a practical balance. It captures the essential risk features of autocallables, particularly exposure to tail events, while remaining computationally feasible and well-suited for pricing in a multi-asset setting.

In the following, we formally describe the Monte Carlo pricing framework, including the stochastic dynamics of the underlying assets, the calibration of model parameters, the simulation procedure, and the variance reduction techniques employed to improve numerical stability and efficiency.

4.2.1 Monte Carlo

The idea of Monte Carlo simulation for pricing derivatives is quite straightforward. One simulates n price paths of the underlying instruments. For each simulated path, the net present value (NPV) of the autocallable is computed by discounting its payoff, which depends on the specific realization of the underlying asset paths. The estimated price of the product is then obtained by averaging the NPVs across all simulation paths.

Formally, the fair value V_0 of the autocallable at inception is given by the risk-neutral expectation of its discounted payoffs:

$$V_0 = \mathbb{E}^{\mathbb{Q}} \left[\sum_{t_j \leq \tau} e^{-rt_j} \cdot C_j(S_{[0,t_j]}) + e^{-r\tau} \cdot R(S_{[0,\tau]}) \right]. \quad (5)$$

Here, t_j denotes the scheduled coupon observation dates, and τ is the stochastic maturity of the product. That is, the first observation date at which the autocall condition is met, or the final maturity if not called early. The term $C_j(S_{[0,t_j]})$ represents the coupon paid at time t_j , which depends on the full path of the underlying basket up to that date. This path dependence captures not only the current state of the underlying, but also the memory feature of the product—allowing previously missed coupons to be paid if the conditions are met at a later date. The function $R(\tau, S_{[0,\tau]})$ denotes the redemption value at τ , which is also path-dependent and reflects whether a knock-in event has occurred.

Each cash flow is discounted at its corresponding payment date using the risk-free rate r , and the expectation is taken under the risk-neutral measure $\mathbb{Q}$. The entire pricing framework—including Heston model calibration and simulation, and Monte Carlo valuation—is implemented using the `QuantLib` library (Ametrano & Ballabio, 2020).

The principal part of the complexity of this method lies in the modelling of the stochastic processes of the underlying instruments, which we discuss next.

4.2.2 The single-asset Heston model

We first describe single-asset individual stochastic process and then elaborate on the dependence structure between these processes, which is needed in order to incorporate the correlations between the different assets. The individual assets each follow a single-asset stochastic volatility model as in Heston (1993). In this model, the risk-neutral dynamics of the asset price process S_t are described by a geometric Brownian motion as follows:

$$dS(t) = (r(t) - q(t))S(t) dt + \sqrt{\nu(t)}S(t) dW^S(t), \quad (6)$$

where $r(t)$ is the risk free rate at time t , $q(t)$ is the dividend rate and $W^S(t)$ is a Wiener process. Moreover, the volatility $\sqrt{\nu(t)}$ is governed by its own mean-reverting stochastic process. Specifically, this process is a CIR process, as developed by Cox et al. (1985). Mathematically this process is given by

$$d\nu(t) = \kappa(\theta - \nu(t)) dt + \eta\sqrt{\nu(t)} dW^\nu(t). \quad (7)$$

Here, $\kappa > 0$ is the reversion rate, θ is the long-run variance. $\eta > 0$ is the parameter for measuring the volatility of volatility (i.e. the degree to which the variance itself fluctuates). The driving force of the variance is the Wiener process $dW^\nu(t)$. The initial variance $\nu(0)$ is also treated as a parameter. Finally, the two processes are assumed to be correlated as follows:

$$dW^S(t)dW^\nu(t) = \rho dt. \quad (8)$$

This correlation is important to capture the so-called leverage effect, which is the well-known phenomenon where an asset's volatility tends to increase when its price decreases. Consequently, this is then captured by a negative correlation between the asset's price and volatility.

4.2.3 Calibration of the Heston model

To implement this model, it is crucial to calibrate the parameters $\Theta = (\kappa, \theta, \eta, \rho, \nu(0))$ such that the stochastic process accurately reflects the market conditions at the time of

pricing. This is done by finding the parameters that match the implied volatility surface as closely as possible. If the Heston model can reproduce the implied volatility surface, it means the model is capable of generating implied volatilities across various strike prices and maturities that closely match those observed in the market. In this manner the model can capture the volatility smile, which is the empirical observation that deeper out-of-the-money options are priced using significantly higher implied volatilities. This is pertinent for pricing autocallables, as the embedded knock-in put option is deep out-of-the-money. Thus using this model, the simulation accurately reflects the underlying process even in periods of extreme losses.

The optimal parameters are obtained by minimizing the sum of squared errors between the market implied volatilities and the implied volatilities generated by the Heston model across N strikes and M maturities of the implied volatility surface. Formally, the minimization problem is given as

$$\arg \min_{\Theta} \quad Err(\Theta) = \sum_{i=1}^N \sum_{j=1}^M (IV_{market}(K_i, T_j) - IV_{model}(\Theta, K_i, T_j))^2, \quad (9)$$

where $IV_{market}(K_i, T_j)$ is the observed implied volatility for strike K_i and maturity T_j . $IV_{model}(\Theta, K_i, T_j)$ is the IV generated by a Heston model with parameter set Θ . The model IV for a given strike and maturity is calculated by first pricing a plain vanilla option with the same strike and maturity using the semi-analytical pricing formula from [Heston \(1993\)](#). The Black-Scholes implied volatility is then extracted from this calculated option price. The optimization is performed using the differential evolution algorithm implemented in the `scipy.optimize` module ([Virtanen et al., 2020](#)), which provides a robust global search method suitable for non-convex objective functions. Since the error function is often highly non-convex, local optimization algorithms tend to get trapped in local minima, necessitating the use of a global search approach. As parameter bounds for the optimization we follow [Bauer \(2012\)](#) and [Crisóstomo \(2015\)](#) in using $\Theta_{bounds} = [(0, 20), (0, 1), (0, 5), (-1, 1), (0, 1)]$. These bounds are chosen to include all economically viable solutions, and are therefore not limiting to the optimization process.

4.2.4 Parsimonious multi-asset Heston model

Since the payoff of autocallables is based on the worst-performing asset in the underlying basket, modelling each asset individually is insufficient, as it would overlook the dependence structure between the assets. As mentioned previously, using the full multi-asset Heston model would result in the need of estimating a large number of parameters. Not only the asset-asset cross-correlations, but also the asset-volatility and volatility-volatility cross-correlations would need to be estimated. In a five-dimensional setting, this leads to high computational costs and unstable parameter estimates. [Dimitroff et](#)

al. (2011) present a parsimonious version of the multi-asset Heston model which aims to circumvent the issues of high-dimensionality. This model solely uses the asset-asset cross-correlations as input. The reasoning behind this approach is that asset-asset correlations are the only dependence parameters that can be reliably estimated from historical return data. In contrast, cross-correlations involving volatility—such as asset-vol and vol-vol correlations—are difficult to estimate accurately, as instantaneous volatility is not directly observable and cannot be inferred reliably from implied or realised volatility measures. In the parsimonious framework, these volatility correlations are instead incorporated indirectly through the asset return correlations and the leverage effect in each asset's individual stochastic volatility process. For instance, a price movement in one asset can influence the price of another through their asset-asset correlation, which in turn affects the second asset's volatility. In this way, indirect dependencies between volatilities are introduced, even though explicit vol-vol correlations are not modelled.

Now extending the notation of the single-asset Heston model to the multidimensional case, consider a system of d assets with price and volatility processes $S_i(t), \nu_i(t)$ for $i = 1, \dots, d$. Then for each asset i we can write

$$\begin{aligned} \begin{pmatrix} dS_i(t) \\ d\nu_i(t) \end{pmatrix} &= \begin{pmatrix} S_i(t)(r(t) - q(t)) \\ \kappa_i(\theta_i - \nu_i(t)) \end{pmatrix} dt \\ &+ \begin{pmatrix} S_i(t)\sqrt{\nu_i(t)} & 0 \\ 0 & \eta_i\sqrt{\nu_i(t)} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \rho_i & \sqrt{1 - \rho_i^2} \end{pmatrix} \begin{pmatrix} dW_i^S(t) \\ dW_i^\nu(t) \end{pmatrix}, \end{aligned} \quad (10)$$

where, $dW_i^S(t)$ and $dW_i^\nu(t)$ are now independent Wiener processes, as the correlation between the asset price and its volatility is modeled explicitly by ρ_i . The only remaining components to describe are the asset-asset cross-correlations between $dW_i^S(t)$ and $dW_j^S(t)$, the correlations asset-vol between $dW_i^S(t)$ and $dW_j^\nu(t)$ and finally the vol-vol correlations between $dW_i^\nu(t)$ and $dW_j^\nu(t)$. Following Dimitroff et al. (2011), we assume that only the asset-asset Wiener processes are correlated, such that $dW_i^S(t)dW_j^S(t) = \rho_{i,j}dt$, while all cross-correlations involving volatility are set to zero, i.e. $dW_i^S(t)dW_j^\nu(t) = 0$ and $dW_i^\nu(t)dW_j^\nu(t) = 0$ for $i \neq j$.

This implies that all dependence between the volatilities is carried through the asset-asset correlation $\rho_{i,j}$, and subsequently transferred to the volatilities through the leverage parameter ρ_i obtained from the single-asset Heston model. As a result, the instantaneous cross-correlations of the model are as follows:

1. $\frac{dS_i(t)dS_j(t)}{\sqrt{(dS_i(t))^2(dS_j(t))^2}} = \rho_{i,j}$
2. $\frac{dS_i(t)d\nu_j(t)}{\sqrt{(dS_i(t))^2(d\nu_j(t))^2}} = \rho_{i,j}\rho_j$

$$3. \frac{d\nu_i(t)d\nu_j(t)}{\sqrt{(d\nu_i(t))^2(d\nu_j(t))^2}} = \begin{cases} \rho_{i,j}\rho_i\rho_j, & \text{for } i \neq j, \\ 1, & \text{for } i = j. \end{cases}$$

As the leverage parameters ρ_i are already obtained from the single-asset Heston model, this means that the entire dependence structure in the d-dimensional model is described by the asset-asset correlation matrix, which we define as Σ^S , where $(\Sigma^S)_{i,j} = \rho_{i,j}$.

A simplifying assumption in this framework is that these correlations are constant over time. While this assumption reduces model complexity and enhances computational feasibility, it is not fully realistic. In practice, correlations tend to increase during periods of market stress—a well-documented phenomenon in financial markets (Kim & Finger, 2000). This means the model may underestimate joint movements in asset prices during extreme events. As a result, it likely overstates the probability of knock-in events, since higher correlations in downturns would make assets fall more uniformly, reducing the likelihood that one index performs significantly worse than the rest.

Although more flexible alternatives such as stochastic correlation models exist, they are significantly more difficult to calibrate. These models typically rely on calibration to multi-asset options, which are infrequently traded and thus produce very noisy estimates. Consequently, the assumption of constant correlations strikes a pragmatic balance, allowing the model to remain robust and computationally feasible while still capturing the key risk dynamics through the correlation matrix Σ^S and the leverage effects embedded in the Heston framework.

4.2.5 Calibration of the multi-asset Heston model

After obtaining the parameters Θ_i for each underlying instrument using the method described in Section 4.2.3, the only remaining task is the calibration of the asset-asset correlations, Σ^S . This is carried out in the same spirit as the single-asset calibration, namely by identifying the parameters that allow the model to most accurately reflect the observed market characteristics. In the single-asset case the objective was to fit the implied volatility surface. Here, however, our objective is to find the Σ^S that best aligns the model with the empirical correlation matrix Σ_K^{emp} , calculated using Equation 3. Thus we want the correlation matrix implied by the Heston-model to match Σ_K^{emp} as closely as possible. It should be noted, as mentioned in Section 4.1, that Σ_K^{emp} is estimated using returns from non-overlapping trading hours (e.g., between US and Japanese markets), which may lead to an underestimation of true correlations. As a consequence, calibrating the model to Σ_K^{emp} could slightly overstate the knock-in risk. In the results, this effect may translate into a modest overestimation of the strategy's standard deviation. However, this is unlikely to materially impact the overall conclusions, as the influence of asset correlations is relatively minor compared to that of the implied volatility of the individual assets.

Instead of optimizing the entire correlation matrix simultaneously, the problem can be simplified by solving each 2-dimensional sub-problem individually, addressing each pair of assets separately. This simplifies the optimization process, as each sub-problem is only dependent on one parameter (see Section 4.2.4). The optimization objective is then formulated as

$$\min_{\rho} |\mathbb{E} [\hat{\rho}^{\text{emp}}(\rho)] - \rho^{\text{emp}}|, \quad (11)$$

where $\hat{\rho}^{\text{emp}}(\rho)$ denotes the empirical correlation implied by the model for a given value of the instantaneous correlation ρ . To compute this, we begin by simulating joint price paths using the already calibrated single-asset Heston processes. This ensures that each underlying asset reflects its own implied volatility surface. The only remaining unknown in a given two-asset sub-problem is the asset-asset correlation parameter ρ . For each candidate value of ρ , we generate the corresponding joint simulations and compute the empirical correlation of the simulated returns using Equation 3. Averaging the resulting correlations across all simulation runs yields the model-implied correlation $\mathbb{E} [\hat{\rho}^{\text{emp}}(\rho)]$. The optimal value of ρ is then obtained by minimising the difference between this model-implied correlation and the empirical correlation observed in the data, ρ^{emp} . Since each sub-problem involves only one parameter, and the expected model-implied correlation $\mathbb{E} [\hat{\rho}^{\text{emp}}(\rho)]$ increases monotonically with ρ , the optimization problem can be efficiently solved using a simple bisectioning method, following the approach of [Dimitroff et al. \(2011\)](#).

This process is repeated $d(d-1)/2$ times to compute the correlation $\rho_{i,j}$ for each pair of assets in the underlying basket of the autocallable product. The next step is to verify that the resulting correlation matrix is positive semi-definite. If it is, the calibration of the model is complete. Otherwise, regularization is applied to the candidate matrix to ensure it becomes positive semi-definite. To achieve this, we employ the regularization method proposed by [Jäckel \(2002\)](#). In short, this technique involves decomposing the correlation matrix into its eigenvalues and eigenvectors, replacing any negative eigenvalues with zero, and then reconstructing the matrix using the modified eigenvalues and the original eigenvectors.

4.2.6 Simulation algorithm

Once the model is fully calibrated, we need to simulate the price paths of the underlying assets in order to price the autocallable. This is done using the Euler discretization of the stochastic processes. Specifically, using the full-truncation scheme of [Lord et al. \(2010\)](#) to avoid issues where the volatility becomes negative. The strength of this scheme lies in its simplicity, while maintaining low bias for short time steps, which is particularly relevant in our case as we use daily time steps. For the time step $t \rightarrow t + \Delta t$, this scheme is given

by

$$S(t + \Delta t) = S(t) + S(t)((r(t) - q(t))\Delta t + S(t)\sqrt{\nu(t)\Delta t}\epsilon^S) \quad (12)$$

and

$$\tilde{\nu}(t + \Delta t) = \tilde{\nu}(t) + \kappa(\theta - \tilde{\nu}^+(t))\Delta t + \eta\sqrt{\tilde{\nu}^+(t)\Delta t}\epsilon^\nu, \quad (13)$$

with

$$\nu(t) = \tilde{\nu}^+(t). \quad (14)$$

Here $(\epsilon^S, \epsilon^\nu)$ follow a bivariate standard normal distribution with correlation ρ . Furthermore we use the notation $(x)^+ = \max(x, 0)$. Note here that the process $\tilde{\nu}(t)$ can become negative, but the variance used for the asset process in Equation 12 is always non-negative due to the truncation.

Of course, for the multivariate Heston-model we need to simulate all underlying assets simultaneously with the correct correlation structure. This is done in the following way:

1. The error terms for the asset price processes $\epsilon^S = (\epsilon_1^S, \epsilon_2^S, \dots, \epsilon_d^S)$ are drawn from a multivariate standard normal distribution with correlation matrix $\Sigma^S = (\rho_{i,j})_{1 \leq i,j \leq d}$.
2. The error terms for the variance processes, $\epsilon^\nu = (\epsilon_1^\nu, \epsilon_2^\nu, \dots, \epsilon_d^\nu)$, are constructed to have a correlation ρ_i with the corresponding asset error term. This is done by first generating d independent standard normally distributed variables $z_1, z_2, \dots, z_d$. Then we calculate $\epsilon_i^\nu = \rho_i \epsilon_i^S + \sqrt{1 - \rho_i^2} z_i$ for $i = 1, 2, \dots, d$.

For each pricing calculation, we simulate 10,000 independent paths of the entire underlying basket, ensuring sufficient convergence of the Monte Carlo estimator, while maintaining computational feasibility.

4.2.7 Variance reduction

As previously mentioned, autocallables are highly path-dependent due to several factors such as the possibility of early redemption, the use of knock-in options and conditional coupon payments. The result of this is that Monte Carlo simulations can be computationally expensive and be subject to high variance in the results, leading to inaccurate pricing unless a large number of simulation paths is used. Thus it is crucial to employ variance reduction techniques to improve the accuracy of the model while decreasing its computational burden. This is done using low-discrepancy quasi-random numbers.

Low-discrepancy quasi-random numbers are deterministic sequences of numbers designed to uniformly cover a multi-dimensional space more effectively than traditional pseudo-random numbers. Pseudo-random numbers are often clustered unevenly, meaning

some regions of the space might be over-sampled, while others are under-sampled. In contrast, Low-discrepancy sequences aim to fit the desired distribution more evenly. The result of this is that the standard deviation of each price calculation is reduced, meaning that the price converges to its expected value with less simulation paths.

In this research we use Sobol sequences, first introduced by [Sobol \(1967\)](#). This method is preferred over other methods such as Faure or Halton sequences due to its superior performance generating low-discrepancy points in multi-dimensional settings, as shown by [Boyle et al. \(1997\)](#). This is an especially important characteristic, because the dimensionality of pricing autocallables is $2d$, as both the asset and volatility process need to be simulated for each underlying instrument. Additionally, Sobol sequences are more efficient computationally, offering faster generation of error terms. See [Figure 5](#) for a comparison of the effectiveness of Sobol sequences versus pseudo-random numbers in generating standard normally distributed values.

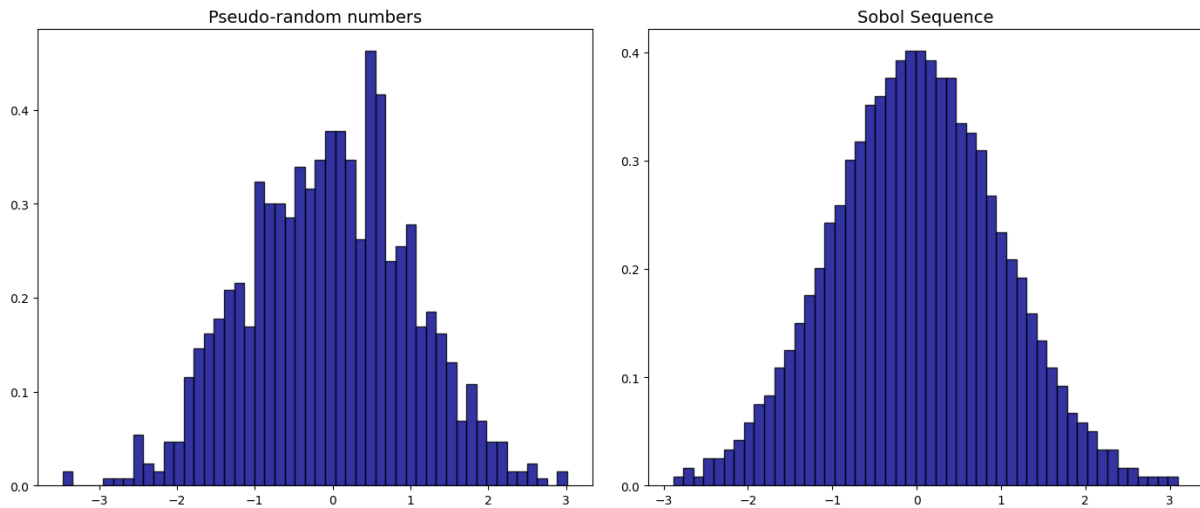


Figure 5: Empirical densities of pseudo-random and Sobol generated standard normal distributions for $n = 1000$

4.3 Constructing the autocallable strategy

Now that we have a method for estimating the coupon rate of an autocallable product and can value these products at any point during their lifetime, the final step is to construct a portfolio in which an investor can remain invested over an extended time period. This involves two key decisions: selecting the specifications of the autocallable product itself, and designing a reinvestment strategy that combines multiple products into a rolling portfolio.

4.3.1 Product specification

The autocallable product is based on a basket of five equity indices, consistent with what we find in our dataset. To determine which indices to include, we prioritize those that appear frequently in the dataset and for which sufficient market data is available (e.g. risk-free rates, implied volatility). As shown in Figure 3a, the five most commonly occurring indices are: Eurostoxx 50, Nikkei 225, Russell 2000, FTSEMIB, and FTSE 100. However, the FTSEMIB's volatility benchmark is only available from 2010 onward, which would exclude several years of valuable data, including the 2008 financial crisis. To maintain a longer historical window and ensure data availability, we replace FTSEMIB with the S&P 500. This index also provides the added benefit of access to its actual implied volatility surface. The final basket therefore consists of: Eurostoxx 50, Nikkei 225, Russell 2000, FTSE 100, and S&P 500.

The other product features were chosen based on the most frequent specifications observed in the dataset. This is to ensure that estimated coupons remain as representative as possible of what would actually be observed in the market. These specifications are as follows:

- A maximum maturity of one year.
- Monthly observation dates for both coupons and autocalls.
- An American-style knock-in, with the barrier set at 60% of the initial level of the underlying indices.
- A coupon barrier at 70% of the initial level.
- An autocall barrier that starts at 99% in month one and steps down by 1% each month thereafter.
- A worst-of payoff structure.
- A memory coupon feature.

4.3.2 Aggregation into a portfolio

While an individual autocallable product has a defined start and end date, our goal is to construct a continuous investment strategy that remains active over a long time horizon. One straightforward approach would be to invest in a new autocallable when the previous one matures. However, this method is vulnerable to idiosyncratic risks associated with specific market conditions at the time of reinvestment.

To mitigate this, we utilise a staggered investment approach that spreads the investments across multiple products with differing initiation dates. This method introduces a form of time diversification, similar to dollar-cost averaging.

The portfolio is initialized during a burn-in period, during which we gradually invest capital into new products. Given the one-month frequency of coupon and autocall observations, we set the burn-in period to four weeks. Each week, 25% of the available capital is invested into a new autocallable. After this initial period, the portfolio holds four active products with staggered issue dates, at which point the formal analysis period begins.

From this point forward, whenever a product matures (either via autocall or at maturity) or a coupon is paid, the proceeds are immediately reinvested into a new product. This ensures the portfolio remains fully invested at all times. Since each product enters the portfolio at a different point in time, their barriers are set relative to different initial levels. This reduces the likelihood that all products will experience adverse events simultaneously, enhancing the overall stability of the strategy.

4.4 Benchmark

The principal aim of this research is to evaluate the long-term risk and return characteristics of a portfolio of autocallable products. A natural benchmark to compare this with would be a portfolio of the underlying indices on which the autocallables are based. [DeMiguel et al. \(2009\)](#) find that equally weighted portfolios often outperform more sophisticated allocation strategies. This stems from the fact that estimation error of the moments of returns often deteriorates the performance of complex strategies. Thus using a 'naive' approach, i.e. assigning equal weights to each asset, generally performs well by circumventing the estimation issue. Following this, our benchmark is an equally weighted portfolio of the underlying indices of the autocallables mentioned in [Section 4.3.1](#).

4.5 Performance evaluation

In this section we discuss the manner in which we evaluate the performance of the autocallable strategy in relation to the benchmark. Firstly, we will evaluate the performance in terms of Sharpe ratio, which is the most common evaluation metric in the field of portfolio management, as it provides a measure of the risk-adjusted performance of the strategy. The Sharpe ratio is calculated using log-returns. Excess returns are then defined as $r_t^{excess} = \ln(\frac{P_t}{P_{t-1}}) - r_f$, where P_t is the price level at time t and r_f is the risk free rate. Subsequently, the Sharpe ratio is calculated as follows:

$$\text{Sharpe Ratio} = \frac{\mathbb{E}[r_t^{excess}]}{\sigma(r_t^{excess})}. \quad (15)$$

For the autocallable products, we use the 3-month EURIBOR rate as the risk-free rate since most of the structures in our dataset are denominated in euros. For individual indices, we use the risk-free rate corresponding to the currency in which the index is based. Finally, for the equally weighted portfolio, we compute the risk-free rate as

the average of the risk-free rates of the indices included in the portfolio. To assess the statistical significance of differences in Sharpe ratios, we follow the approach of Kirby & Ostdiek (2012). In short, they apply the Generalized Method of Moments (GMM) to derive asymptotic standard errors of the difference in Sharpe, while accounting for heteroskedasticity and autocorrelation. Since the distribution of the test statistic may deviate from a normal distribution, they further employ the stationary block bootstrap of Politis & Romano (1994) to obtain empirical p-values, ensuring robust inference without relying on normality assumptions.

Solely looking at the annual return, standard deviation, and Sharpe ratio does not however capture the full risk profile of autocallables. The primary goal of autocallable products is to deliver predictable returns under most market conditions. However, for investors seeking such predictability, it is crucial to understand the risk arising from the written knock-in put option embedded in these products, since a knock-in event can lead to relatively large losses. This information is key for investors in determining whether the enhanced yields of autocallables adequately compensate for this risk.

A valuable metric for this purpose is expected shortfall (ES), which measures the average loss in the worst-case scenarios beyond a specified confidence level. A common approach to estimate ES is via Monte Carlo simulation, such as in Tzeng et al. (2018) and Hsieh et al. (2014). We generate a large number of random scenarios for portfolio returns using the same data-generating process as employed for pricing the autocallables, namely the multi-asset Heston model. At each point in time, we simulate possible price paths over a one-year horizon, corresponding to the maximum maturity of the autocallable products. After calculating the resulting losses at the end of this period and sorting them in ascending order, we identify the Value at Risk (VaR) at a chosen confidence level α (commonly 5%), which represents the loss threshold such that only the worst $\alpha\%$ of outcomes exceed it. The expected shortfall is then computed as the average loss among all outcomes that exceed this VaR. Formally, if VaR_α is the VaR at confidence level α , the expected shortfall is given by:

$$\text{ES}_\alpha = \mathbb{E}[L \mid L > \text{VaR}_\alpha], \quad (16)$$

where L denotes the loss. This measure provides a clear indication of the downside risk associated with the autocallable strategy, thereby enabling a more comprehensive risk-adjusted evaluation.

5 Results

In this section, we present the results of our analysis, offering a comprehensive assessment of what an investor can expect from a rolling portfolio of autocallable structured products

and how its performance compares to the underlying indices. First, we examine the general performance of the analysed strategies in terms of their risk and return characteristics. We then investigate the key drivers behind the observed performance differences. Following this, we conduct a more in-depth risk assessment by analyzing the expected shortfall of the different strategies over time. Lastly, we analyse the evolution of the coupon rates of the autocallable portfolio to give a more complete picture of the compensation investors receive for taking on the associated risks.

5.1 Performance comparison

Figure 6 illustrates the monthly price evolution of the autocallable strategy, the equally weighted portfolio of the underlying indices. For reference, each individual index is included as well as the nominal value of the autocallable portfolio. The nominal value is included to highlight the strategy’s steady growth under normal conditions and to clearly indicate when products expire with a knock-in, in which case they return less than their nominal value. This is the only case where the nominal value of the portfolio can decrease. In the case where the worst performing index falls below the coupon barrier, the coupon payment is withheld, resulting in a mostly flat nominal value of the portfolio. A monthly timeframe was chosen because it provides a clearer, less noisy view of performance compared to a weekly timeframe, without omitting important data. For comparisons using other timeframes, see Appendix A. The data span from 5 May 2006, marking the end of the portfolio’s burn-in period, to 31 August 2023.

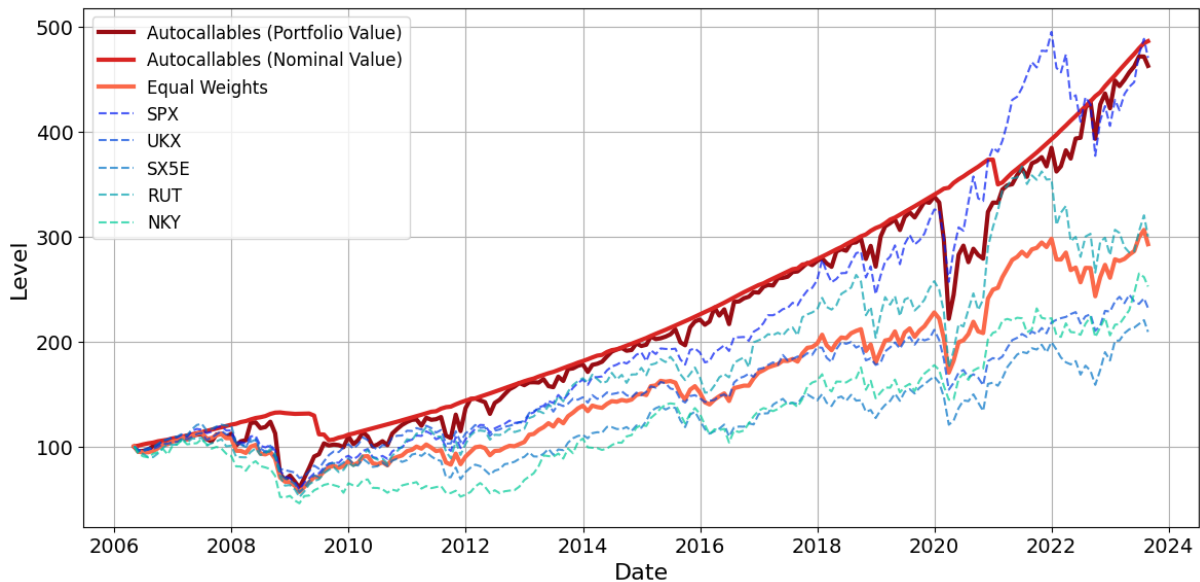


Figure 6: Comparison of the monthly price evolution of the autocallable strategy with the equally weighted portfolio of the underlying indices. For reference, each underlying index is also shown individually (including dividends).

The indices and the equally weighted portfolio start at a reference value of 100, and the autocallable strategy is initiated at a nominal value of 100 (corresponding to an investment of 100), although the pricing model values it at 99.68 at this time. A more detailed discussion of the pricing of the structures will be provided later in this section.

During the analysis period, the autocallable strategy’s performance remains consistently above the equally weighted benchmark and surpasses most of the individual indices in terms of overall returns. Notably, only the S&P 500 is on par with the autocallable strategy. These results suggest that, although both strategies rely on the same underlying instruments, the autocallable structure can deliver superior returns compared to a simple equal-weight approach. However, to determine whether this reflects genuine outperformance on a risk-adjusted basis—rather than merely higher risk-taking—we next compare each strategy’s returns relative to its volatility.

Table 1 presents the annualised return, standard deviation, and Sharpe ratio of all investigated portfolios, calculated over different timeframes to provide a more nuanced understanding of their performance. These metrics are computed using log-returns to ensure consistency across timeframes. This is done because of the additivity of log-returns, ensuring that, for example, the annualised return derived from weekly observations matches the return observed over a full year. From here on we simply write ‘returns’ to denote log-returns. For longer timeframes (monthly, quarterly, and yearly), returns were obtained by summing the weekly returns within each respective period. An adjustment was necessary for the first and last observation periods, as the start and end dates of the analysis did not align perfectly with full quarters or years. To address this, missing periods were imputed using the average return over the entire analysis period, ensuring greater comparability across different timeframes.

Table 1: Performance comparison between the autocallable portfolio, the equally weighted portfolio, and the underlying indices. The Sharpe ratio, annualized returns and annualized standard deviation are presented for different timeframes.

Time Frame	Metric	Benchmark Comparison		Individual Indices				
		Autocallables	Equal Weights	SPX	UKX	SX5E	RUT	NKY
Weekly	Sharpe ratio	0.303	0.266	0.400	0.184	0.151	0.200	0.230
	Annualized return	7.93	4.98	7.38	3.32	3.34	4.79	5.06
	Standard deviation	26.18	18.70	18.45	18.04	22.07	23.95	22.01
Monthly	Sharpe ratio	0.372	0.302	0.446	0.230	0.179	0.222	0.248
	Annualized return	7.95	4.99	7.39	3.32	3.35	4.79	5.07
	Standard deviation	21.38	16.54	16.57	14.48	18.74	21.55	20.48
Quarterly	Sharpe ratio	0.367	0.282	0.426	0.216	0.170	0.209	0.252
	Annualized return	7.94	4.97	7.36	3.30	3.33	4.76	5.08
	Standard deviation	21.65	17.63	17.27	15.23	19.60	22.78	20.17
Yearly	Sharpe ratio	0.466	0.262	0.365	0.202	0.159	0.228	0.236
	Annualized return	7.87	4.89	7.26	3.22	3.27	4.67	5.08
	Standard deviation	16.89	18.64	19.91	15.90	20.52	20.51	21.57

Looking at the weekly timeframe, the autocallable strategy achieves an annualised return of 7.93% compared to 4.98% for the equally weighted benchmark. However, this higher return comes with greater variability, as evidenced by a standard deviation of 26.18 versus 18.70 for the benchmark. Notably, the autocallables exhibit a higher Sharpe ratio (0.303) compared to the benchmark's 0.266, indicating better risk-adjusted performance despite the elevated volatility. It is also worth mentioning that, on a weekly basis, the standard deviation of the autocallable strategy is the highest among the individual indices.

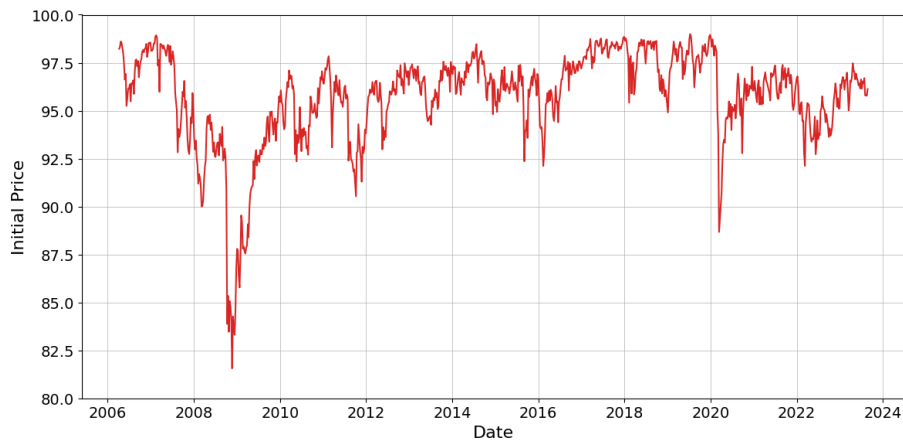
When we shift to monthly and quarterly timeframes, the picture changes. Although all portfolios experience a reduction in volatility when moving from weekly to longer periods, the decline is most pronounced for the autocallable strategy. Its standard deviation drops to 21.38 (monthly) and 21.65 (quarterly), a notable decrease from the 26.18 observed weekly. In contrast, the benchmark's volatility decreases moderately from 18.70 weekly to 16.54 monthly and 17.63 quarterly. Moreover, on these timeframes the autocallables are no longer the most volatile; for instance, the Russell 2000—a traditionally higher-variance index due to its small-cap composition—shows volatilities of 21.55 and 22.78, respectively. This relative alignment is also reflected in the Sharpe ratios, with the monthly Sharpe ratio rising to 0.372 for the autocallables versus 0.302 for the benchmark, and 0.367 versus 0.282 on a quarterly basis. The differences in risk-adjusted performance, therefore, become more pronounced at these longer frequencies, increasing from a difference of 0.027 at the weekly level to 0.070 and 0.085 on the monthly and quarterly timeframes, respectively.

For the yearly timeframe, this trend continues. The volatility of the autocallable portfolio falls to 16.89, compared to 18.64 for the benchmark, suggesting that over longer periods, the autocallable strategy delivers more predictable returns. This is attributable to the strategy's tendency to follow the steady growth of its nominal value under normal market conditions, coupled with its strong bounce-back potential during crises—evidenced during events such as the 2008 Global Financial Crisis and the COVID-19 pandemic. Consequently, the annual Sharpe ratio for the autocallables is 0.466, substantially higher than the 0.262 for the benchmark, a difference of 0.204. In fact, on an annual basis, the autocallable strategy even outperforms the S&P 500, the strongest performer among the individual indices. It should be noted, however, that the yearly analysis is based on only 18 observations, which may affect the reliability of the standard deviation estimate, though it still provides valuable insights into the performance dynamics of the autocallable structured products.

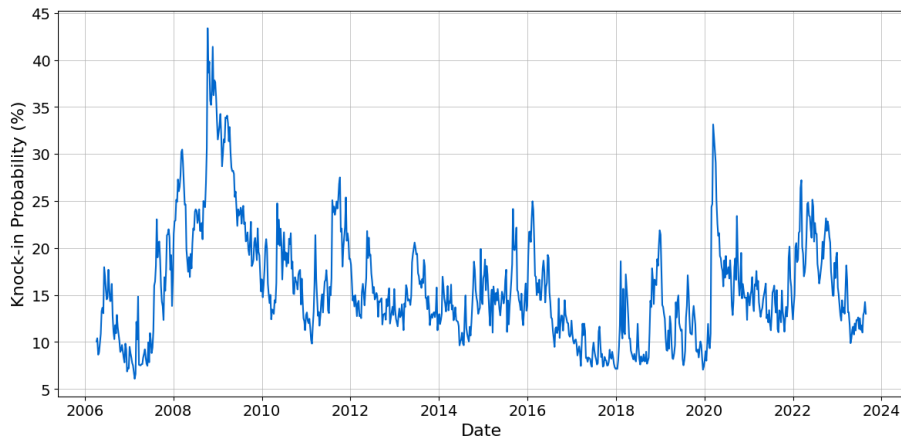
Using the significance test described in Section 4.5, we find that the difference in Sharpe ratios between the autocallable strategy and the benchmark is not statistically significant across all time frames, meaning we fail to reject the null hypothesis of equal risk-adjusted performance. The p-values of the test are 0.387, 0.347, 0.300, and 0.203 for return frequencies ranging from weekly to yearly. Although the p-values decline with lower return frequency, they all remain above conventional significance thresholds of 5% and 10%.

However, it is well-documented that Sharpe ratio comparisons require very long sample periods to achieve sufficient statistical power, especially when differences are moderate (e.g. Ledoit & Wolf (2008), Lo (2002)). Given these limitations, it remains essential to interpret the findings within their economic context, as statistical insignificance does not necessarily imply that the results are not relevant practically speaking.

The differences in volatility of the autocallable portfolio across different timeframes can be explained by examining the pricing of the autocallables in more detail. Figure 7 shows the initial pricing of newly issued autocallables over time (where 100 signifies par, meaning at the nominal value), as well as the knock-in probabilities of those products as estimated by the pricing model (calculated as the percentage of simulation paths where a knock-in occurred).



(a) Initial prices.



(b) Model-implied knock-in probabilities.

Figure 7: Model-implied initial prices and knock-in probabilities for newly issued autocallables.

Our analysis shows that the initial prices of autocallable products are consistently below par for every observation date. On average, the initial price is 95.72, indicating that these products are overpriced by 4.47% relative to their fair value as determined by

the pricing model. This finding aligns with the literature discussed in Section 1, which suggests that structured products generally tend to be issued at a premium.

Additionally, the initial price fluctuates significantly over time. The most extreme case occurred during the 2008 financial crisis when the initial price dropped to as low as 81.56. This occurred, despite a substantial increase in coupon rates during this period of high market stress. For instance, even the coupon rate of 38.3% was insufficient to offset the heightened risk of a knock-in event. This general pattern is evident throughout the analysis period: higher risk, as reflected in the model-implied knock-in probability, strongly correlates with a lower initial price. This suggests that in riskier market conditions, issuers demand higher margins to compensate for the increased uncertainty.

The variability in initial prices has a significant impact on the valuation of an autocallable portfolio. When an autocallable product matures, a new one is added to the portfolio at par. However, when evaluating the portfolio's fair value, the pricing model is used, which typically values the newly added product below par. This systematic discrepancy—where the purchase price exceeds the model-implied value—causes an immediate drop in the portfolio's fair value upon reinvestment. As a result, this reinvestment dynamic introduces considerable short-term fluctuations in the portfolio's value. This explains why the standard deviation of the autocallable portfolio appears uncharacteristically high when measured on a weekly basis, and decreases significantly when observed over longer observation intervals.

5.2 Autocallable performance: implied vs. realised volatility

The question remains why autocallables are able to achieve a consistently better risk-adjusted performance than the benchmark, especially considering that the autocallable products are on average overpriced as determined by the pricing model. We attribute this primarily to the discrepancy between implied volatility and realised volatility. Autocallables are priced using implied volatility, which tends to be systematically higher than realised volatility. As a result, the probability of knock-in events and breaches of the coupon barrier is often overestimated by the pricing model, leading to an overstatement of downside risk.

To demonstrate this effect, Figure 8 shows the probability of an autocallable product terminating in each month of its lifetime, either through early redemption or a knock-in event. The figure presents the termination months of all products that have matured, so excluding those still active at the end of the analysis period. It also displays the termination probabilities predicted by the pricing model, allowing for a direct comparison between realised and model-implied outcomes. The model-implied termination probabilities are obtained by first aggregating the termination months across all simulation iterations for each observation date and then normalizing by the total number of observations.

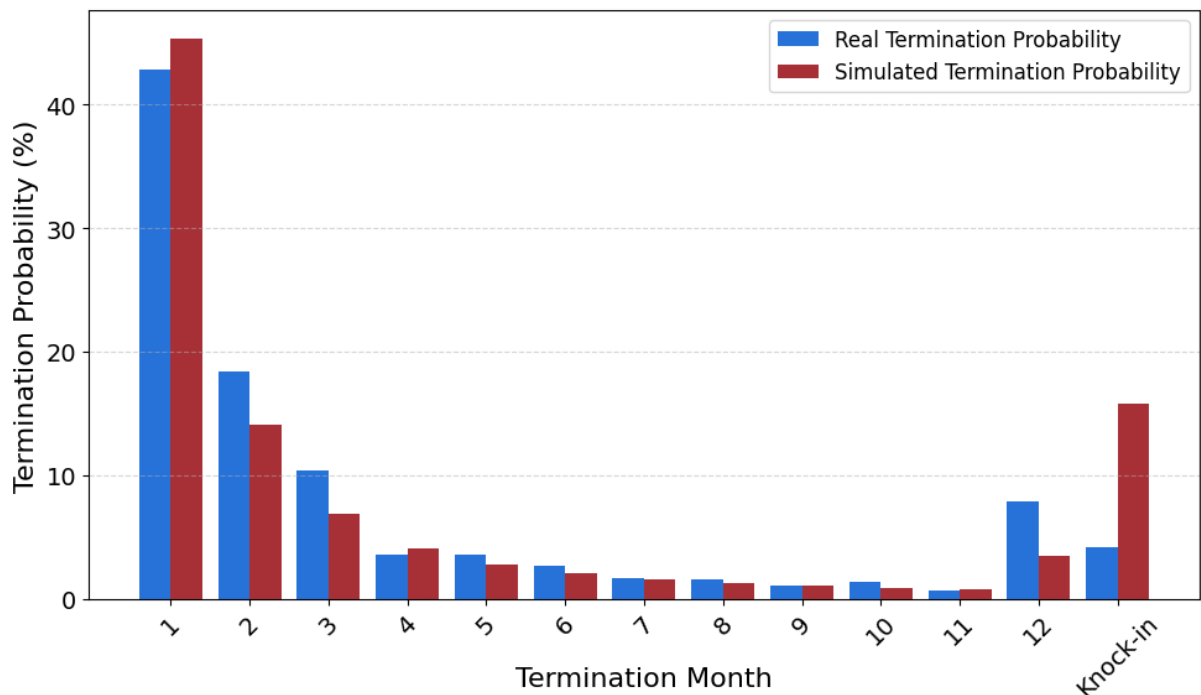


Figure 8: Comparison of the realized termination month distribution and the model-implied termination probabilities across each month in an autocallable’s lifetime.

For the most part, Figure 8 shows that the distribution of realised and model-implied termination months aligns well. In both cases, the most common outcome for an autocallable is expiration at the first autocall date, after which the probability declines rapidly for the next few months and then gradually approaches zero. The least likely termination month is the 11th, with only 0.79% of simulations and 0.68% of actual autocallables expiring at that point in the product’s lifetime.

However, material differences emerge in the final two termination categories: expiration in the 12th month and expiration with a knock-in. The probability of expiring in the 12th month increases relative to the previous months, because this includes cases where the worst-performing index never recovers above the autocall barrier but is forced to terminate at maximum maturity. While knock-in cases also expire in the 12th month, they are categorized separately in this graph, to highlight the differences between the two outcomes. The key discrepancy is that the realised probability of expiring in the 12th month (7.91%) is more than twice the model-implied probability (3.48%), indicating that the pricing model underestimates this outcome.

An even larger discrepancy appears in expirations where a knock-in event has occurred. The model assigns a 15.75% probability to knock-in events, whereas in reality, only 4.18% of autocallables expired this way. This is close to four times lower than predicted. An implication of this is that the deep out-of-the-money implied volatility used in pricing significantly overestimates the likelihood of a knock-in. As a result, knock-ins occur far less frequently than expected, allowing investors to collect higher coupons consistently,

since coupon rates are determined by the level of the implied volatility surface. This is the largest contributing factor to the outperformance of the autocallable strategy compared to the underlying indices.

This finding supports existing research showing that implied volatility tends to be higher than realised volatility, as discussed in Section 1. By selling volatility, which autocallables do by embedding a written knock-in put option, investors can systematically capture this difference, commonly referred to as the volatility risk premium (VRP).

Of course, volatility selling can be achieved through other instruments, such as plain-vanilla options or VIX futures, which has been shown to enhance risk-adjusted performance in previous studies (e.g. [Ge \(2016\)](#), [Israelov et al. \(2017\)](#)). We argue, however, that autocallables can provide a particularly effective way to capture the volatility risk premium. This is because their early redemption feature allows for repeated reinvestment, enabling investors to capture the VRP multiple times within the same investment horizon.

In a standard put-selling strategy, the VRP is collected only once per contract cycle, and the investor must wait until expiration before reinitiating the position. In contrast, an autocallable product may redeem early if market conditions are favourable, freeing up capital to be reinvested into a new autocallable. This creates a compounding effect, where investors systematically harvest VRP at a higher frequency compared to traditional volatility-selling approaches. Additionally, because the initial pricing of the autocallable factors in the full downside risk over the maximum possible lifetime, the VRP embedded in the structure is likely higher than what would be implied for a short-dated put. For these reasons, a rolling portfolio of autocallable products is a particularly effective method for capturing the volatility risk premium efficiently.

5.3 Expected shortfall comparison

While we have examined the risk-adjusted performance of the autocallable strategy relative to the benchmark, it is also important to analyse the tail risk of both strategies. Extreme losses, though rare, are a critical consideration for investors assessing the viability of a given strategy. Figure 9 presents the expected shortfall (ES) for both the autocallable strategy and the equally weighted benchmark portfolio, expressed as a percentage loss relative to total portfolio value. This analysis is conducted at a 5% significance level, meaning that the expected shortfall represents the average loss in the worst 5% of scenarios.

A key observation is that the expected shortfall of the autocallable portfolio exceeds that of the benchmark in most cases. On average, the ES for the autocallables is 55.81%, compared to 46.16% for the benchmark. This implies that in the worst 5% of scenarios, the autocallable portfolio is expected to experience nearly 10% greater losses than the equally weighted portfolio of the underlying indices.

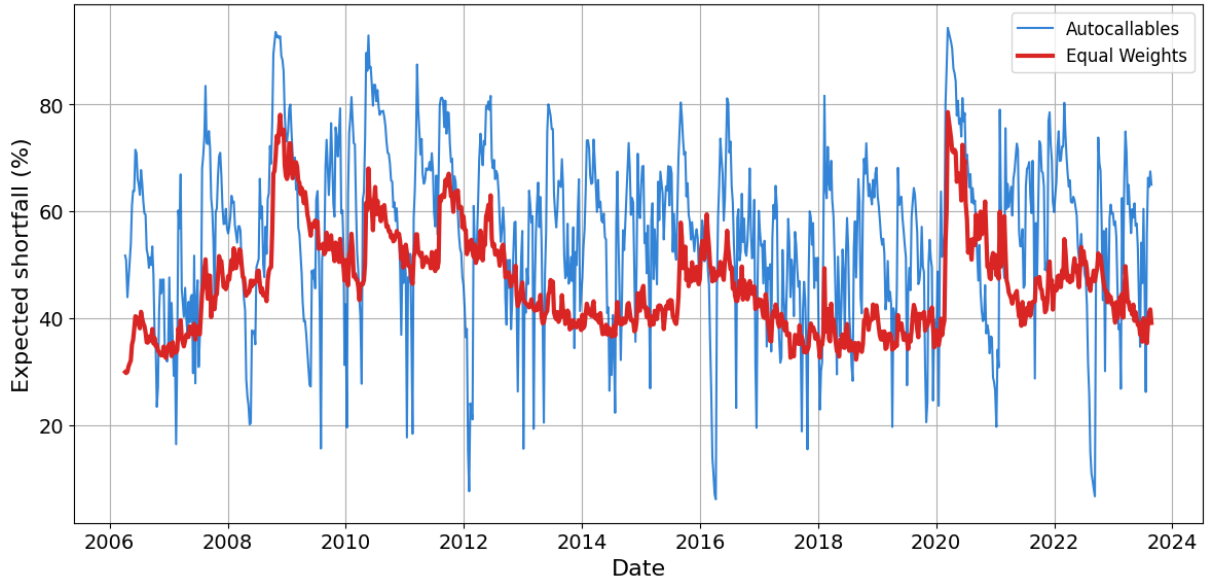


Figure 9: Expected shortfall (5%) comparison between the autocallables portfolio and equally weighted indices.

This difference is primarily driven by the worst-of feature embedded in the autocallable structure. Since ES is computed over the most extreme loss scenarios, it overwhelmingly consists of instances where the knock-in barrier has been breached. Given that the model-estimated knock-in probability is at least 6.09% over the analysis period for newly issued autocallables, a substantial portion of these worst-case scenarios involve knock-ins. In these cases, the autocallable’s payoff is strictly worse than that of the benchmark because the benchmark reflects the average performance of the underlying indices, whereas the autocallable payoff is determined by the worst-performing index.

Nevertheless, there are specific scenarios where the expected shortfall of the autocallable portfolio is lower than that of the benchmark. This occurs, for instance, when some autocallables in the portfolio have a worst-performing index that remains comfortably above the autocall barrier as the next observation date approaches. In these cases, early redemption becomes highly likely, ensuring both the principal and a coupon payment, since the risk of any index breaching the autocall barrier within the remaining time is minimal. As a result, the portfolio’s expected shortfall drops significantly during such periods.

These factors—the nuanced behaviour of autocallables around observation dates and the worst-of feature—also explain why the expected shortfall of the autocallable portfolio varies significantly more than that of the benchmark.

The comparison of expected shortfall between the autocallable portfolio and the benchmark carries important implications for investors evaluating these products. While autocallables provide stable returns under most market conditions, they are more exposed to extreme downside risks compared to a diversified investment in the underlying indices. The worst-

of feature amplifies losses in the most severe market scenarios, making the tail risk of autocallables greater than that of the benchmark portfolio.

Moreover, the higher variability in expected shortfall over time suggests that investing in autocallables requires more sophisticated risk management. Unlike a traditional investment, where downside risk tends to follow a more predictable pattern, the risk profile of an autocallable portfolio can shift abruptly due to its dependence on both early redemption dynamics and the behavior of the worst-performing index. This added complexity means that investors must be prepared for unexpected fluctuations in downside risk, particularly in volatile market environments. Ultimately, investing in a portfolio of autocallables requires balancing the potential for enhanced risk-adjusted performance, driven by efficient harvesting of the volatility risk premium, against the need for disciplined risk management and tolerance for larger losses during black-swan events.

An important final consideration is whether the tail risk of the autocallable portfolio could be mitigated through broader diversification. Expanding the portfolio to include autocallables based on a wider variety of underlyings, such as indices from different regions or entirely different asset classes, could reduce the concentration risk inherent in a portfolio built solely on a fixed set of equity indices. Since the worst-of feature penalises the lowest-performing underlying, greater heterogeneity in the basket of products would reduce the likelihood that many autocallables experience simultaneous knock-in events. This, in turn, could help smooth the payoff distribution over time, leading to a lower and more stable expected shortfall. This underscores the potential benefits of cross-product diversification as a risk management tool in autocallable strategies. However, a detailed investigation of this diversification effect falls outside the scope of this paper and is left for future research.

5.4 Coupon rates

To provide a more comprehensive understanding of how the autocallable portfolio evolves over time, we also examine the coupon rates of the strategy throughout the analysis period. These coupon rates are determined by the regression model discussed in Section 4.1, meaning they are primarily influenced by the implied volatilities of the underlying indices. As a result, coupon rates closely reflect the risk level of an autocallable at any given time. Figure 10 presents the historically estimated coupon rates over time, alongside the average coupon rate of the autocallable portfolio. The average coupon rate of the portfolio is computed as a weighted average, where each autocallable's coupon rate is weighted by its nominal value, reflecting the proportion of the portfolio invested in that product. This metric represents the effective coupon rate received by the portfolio, assuming that all autocallables are above their coupon barrier, which is the case for the vast majority of the time.

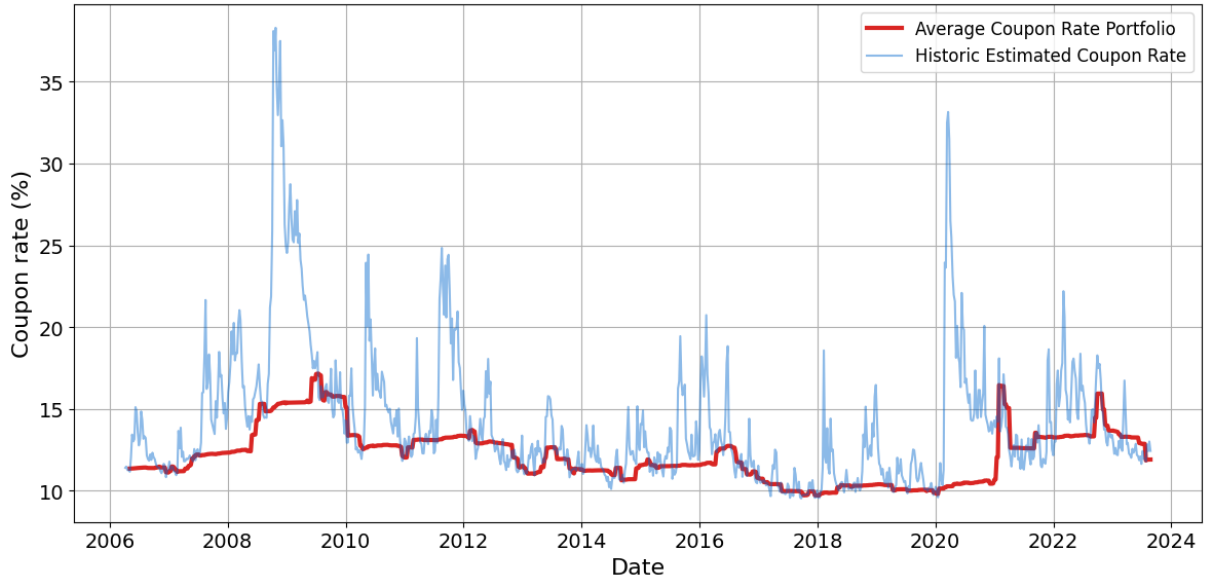


Figure 10: Evolution of the portfolio’s average coupon rate over time, compared to the estimated coupon rate for a newly issued autocallable at each point in time.

The historically estimated coupon rates generally range between 10% and 20% annually, with exceptions occurring during periods of extreme market volatility. The coupon rate has only exceeded 30% twice: during the peaks of the 2008 Global Financial Crisis (GFC) and the COVID-19 crisis. Over the full analysis period, the average coupon rate for newly issued autocallables is 14.10%.

However, when examining the actual realised coupon rate of the autocallable portfolio, we observe a notable discrepancy. The portfolio’s average coupon rate over the analysis period is 12.15%, nearly two percentage points lower than the average historically estimated rate. Additionally, the portfolio coupon rate is much smoother over time and exhibits less extreme values. For instance, the maximum observed portfolio coupon rate is 17.15%, recorded near the end of the Global Financial Crisis, significantly below the peak historical estimates for newly issued products.

The difference between the historically estimated coupon rates and the realised portfolio coupon rates arises from a form of selection bias driven by the early redemption feature of autocallables. During periods of high volatility, newly issued autocallables offer higher coupon rates to compensate investors for the increased risk. However, these spikes in volatility typically follow significant market drawdowns, which drive implied volatility higher. This is a phenomenon linked to the leverage effect, as discussed briefly in Section 4.2.2. As a result, in times of heightened volatility, many existing autocallables are unlikely to redeem early, since their worst-performing index is often well below the autocall barrier. These products remain in the portfolio for an extended period, either until the market recovers (by which point volatility has likely decreased) or until they reach their maximum maturity. This effect was particularly evident at the end of the

2008 financial crisis and in early 2021, roughly a year after the start of the COVID-19 crisis. In both cases, the coupon rate of the entire portfolio increased, but notably not to the same extreme levels observed in newly issued products at the peak of the crises.

Due to this effect, the autocallable portfolio exhibits relatively stable coupon rates, almost always within the 10–15% range. This stability enhances the predictability of returns over time, making the portfolio a consistent income-generating instrument in most market conditions. However, it also imposes constraints on investors pursuing active management, as they cannot freely time their purchases to take advantage of higher coupon rates during periods of elevated volatility. Since the termination of products depends on market conditions, the reinvestment of proceeds is largely dictated by the redemption schedule, rather than by an investor’s own timing. As a result, investors seeking to optimise yield through discretionary adjustments may find limited flexibility in managing an autocallable portfolio, compared to other fixed-income strategies. However, while timing opportunities are constrained, investors still retain control over the structural design of the products in their portfolio. This includes selecting the underlying indices or assets and determining the levels of the various barriers of the product.

6 Conclusion

In this research, we investigated the long-term performance of autocallable structured products and compared it to that of a traditional equity investment in the underlying indices. To enable this comparison, we developed a novel methodology to construct synthetic return data for a portfolio of autocallable products. This involved first extrapolating the available data to estimate plausible coupon rates using observed market variables, and subsequently employing Monte Carlo simulations under a parsimonious multi-asset Heston model to price the products over their lifetime. The resulting strategy, based on continuous reinvestment of autocallables upon product maturity, could then be compared to the equally weighted portfolio of the underlying indices.

The primary goal of this research was to provide investors with a clear understanding of the risk-return characteristics of an autocallable portfolio. Given the inherent complexity of these products, it is particularly important for investors to gain a realistic understanding of their performance and associated risks so they can make well-informed decisions.

Our results show that the autocallable strategy consistently outperforms the benchmark on a risk-adjusted basis. This outperformance is most pronounced at lower observation frequencies (i.e. monthly, quarterly, and yearly), where the effect of short-term pricing deviations diminishes and the strategy more closely tracks the nominal value of the portfolio. At a weekly frequency, however, the standard deviation is higher due to the pricing model’s deviations from par at product initiation, leading to greater fluctuations in the portfolio value when reinvesting into new autocallables.

A key driver of the strategy’s outperformance is its ability to capture the volatility risk premium, reflected in the persistent gap between implied and realized volatility. Because of the inflated implied volatility, the simulations from the pricing model overestimates the likelihood of knock-in events by a factor of nearly four. Thus one receives a high coupon rate based on this overestimated tail risk, while the actual risk is considerably lower. Moreover, we argue that the early redemption feature of autocallables allows this volatility premium to be harvested more efficiently than in traditional volatility-selling strategies.

However, we also observe that the 5% expected shortfall of the autocallable portfolio is, on average, nearly 10% higher than that of the equally weighted portfolio of underlying indices. We attribute this to the worst-of feature embedded in the autocallables. For investors, this implies a trade-off: while the strategy offers the potential for enhanced risk-adjusted performance through efficient capture of the volatility risk premium, it also carries greater vulnerability to losses in extreme scenarios.

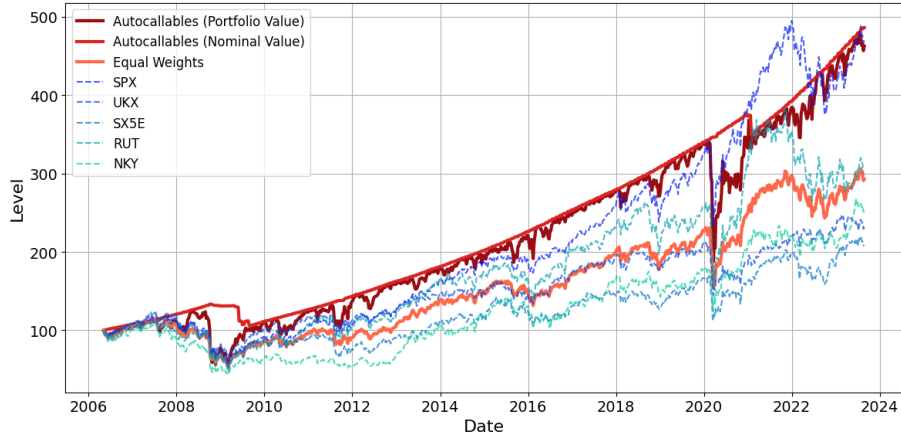
There are several limitations to this research. The first was the unavailability of all necessary data for the implied volatility surfaces. The IVS for three of the five underlying indices was not available, and therefore had to be estimated using the available data. While this does not affect the terminal returns of the autocallable portfolio, it likely distorts the intermediate pricing of the autocallables slightly, which potentially changes the standard deviation of the strategy. It is, however, difficult to say in which direction. Additionally, a longer time period of data would likely grant better statistical power to confidently differentiate between the autocallable strategy and the benchmark.

The second limitation concerns the pricing model, and specifically its use of correlations. We utilise empirical estimates of the correlations instead of implied correlations, and the model assumes these correlations to be constant over time. As empirical correlations are inherently backward-looking, they may not accurately reflect future market dynamics. Moreover, the constant correlation structure ignores some well-documented market dynamics, such as the finding that cross-asset correlations increase during market crashes, which can impact the probability of a knock-in event occurring.

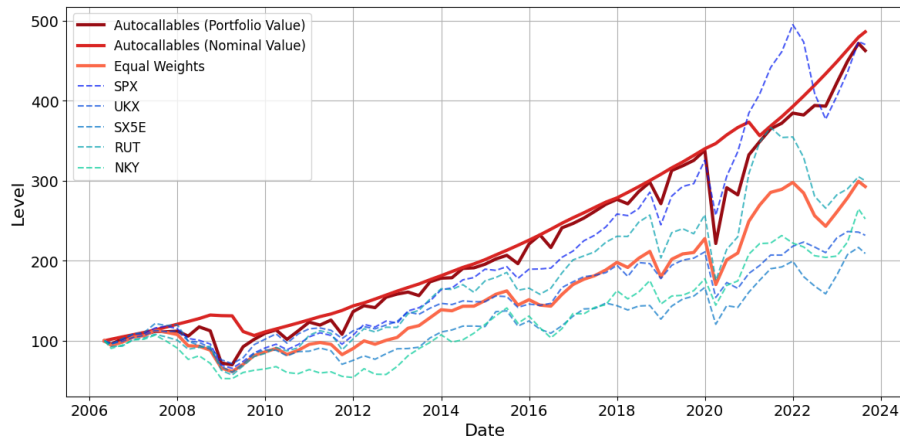
Future research could extend this work by applying the same methodology to a more complete dataset and using a pricing model that incorporates implied correlation surfaces, in order to test the robustness of the findings of this paper. Another natural extension would be to explore dynamic portfolio optimisation, allowing for the selection of different autocallable product features (e.g. barrier levels, time-to-maturity and autocall frequency).

7 Appendix

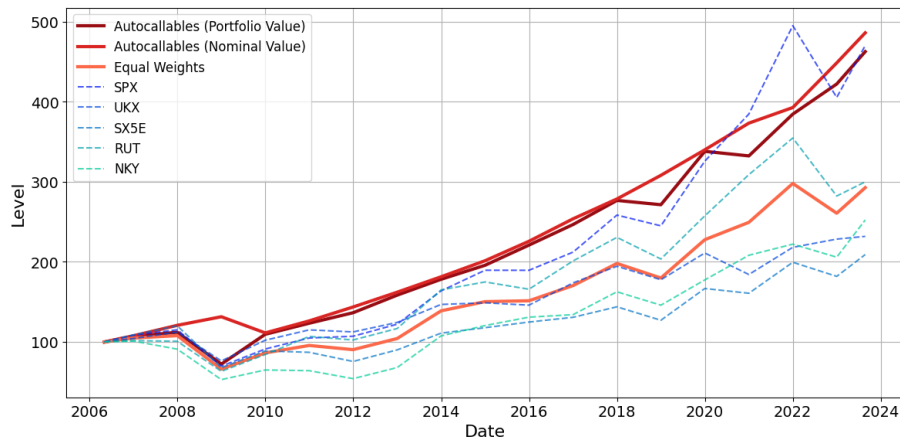
A Tables and Figures



(a) Weekly

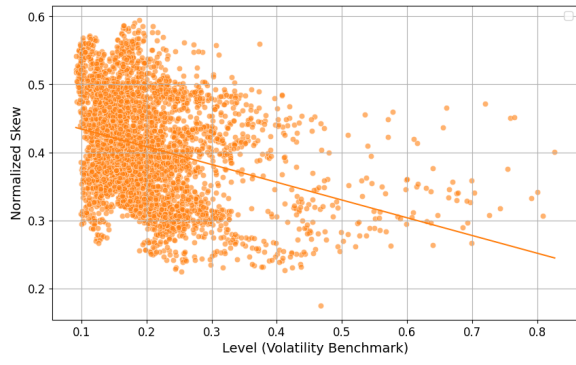


(b) Quarterly

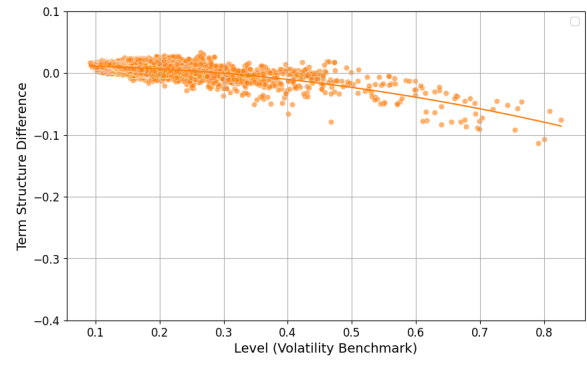


(c) Yearly

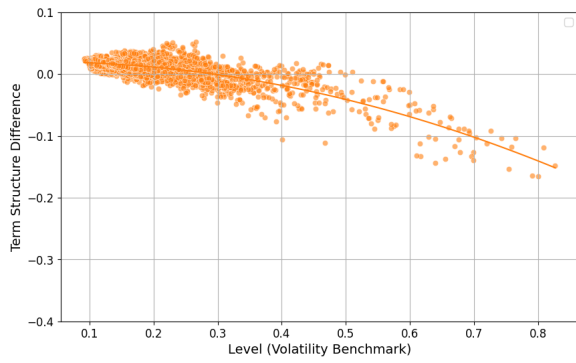
Figure 11: Price evolution of the autocallable and indices portfolios using other time-frames.



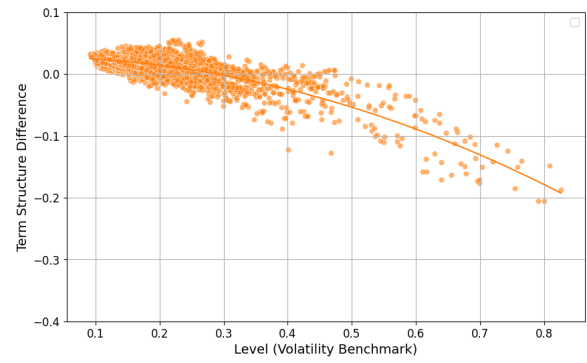
(a) Normalized skew



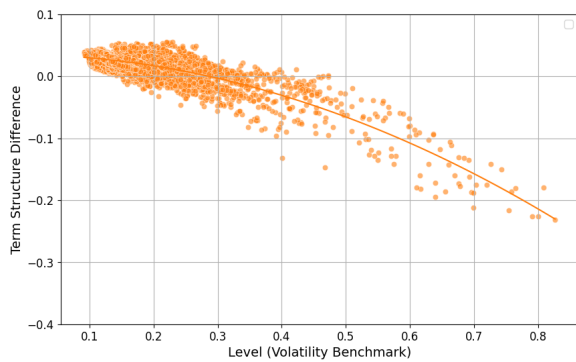
(b) 2M



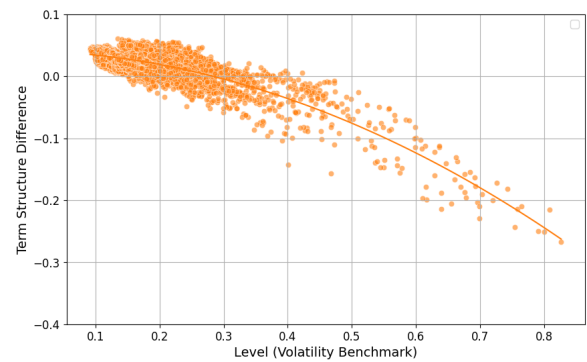
(c) 3M



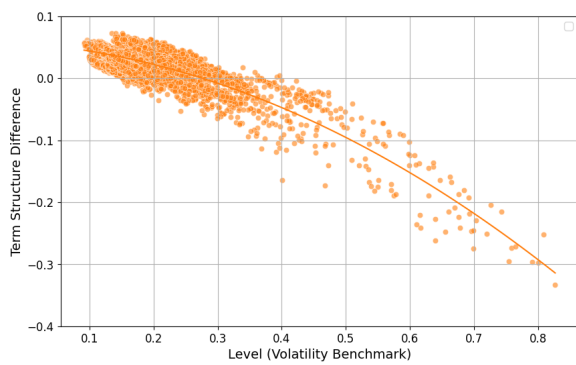
(d) 4M



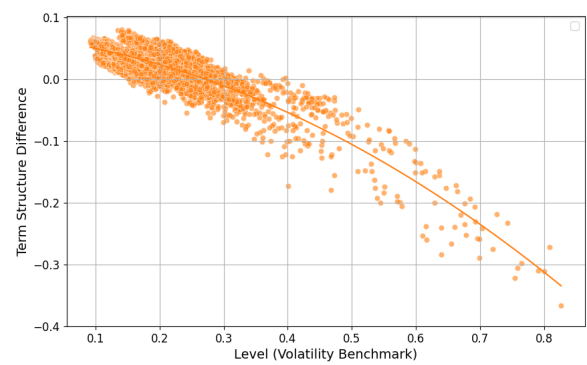
(e) 5M



(f) 6M

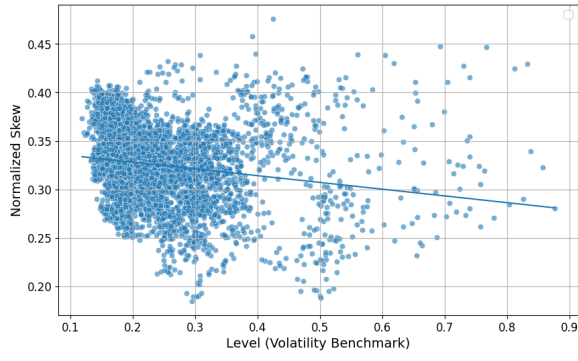


(g) 9M

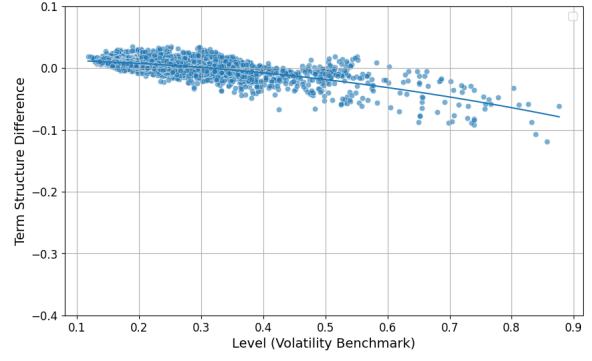


(h) 12M

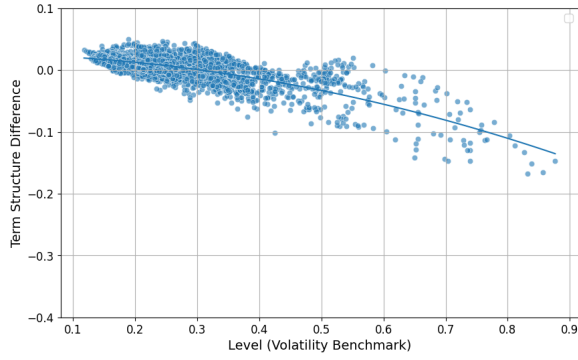
Figure 12: Normalized skew and term structure regressions on the level of the volatility benchmark for the S&P 500.



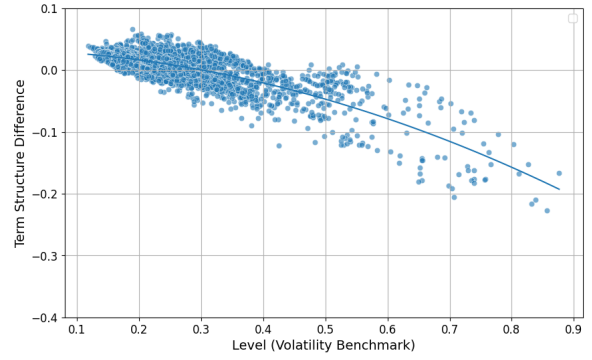
(a) Normalized skew



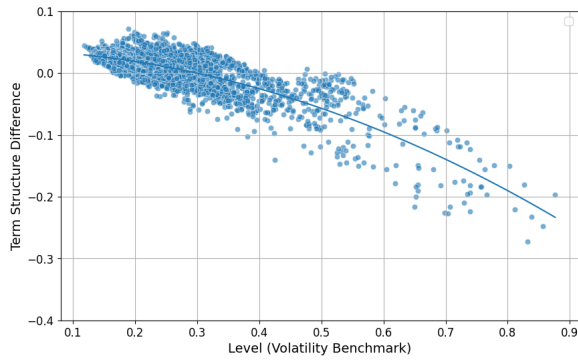
(b) 2M



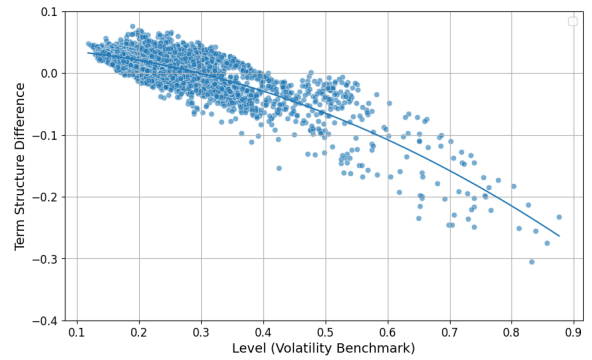
(c) 3M



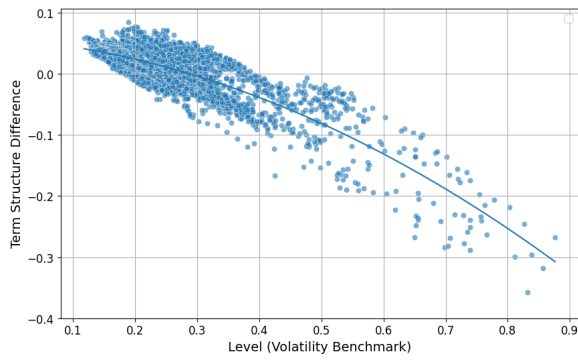
(d) 4M



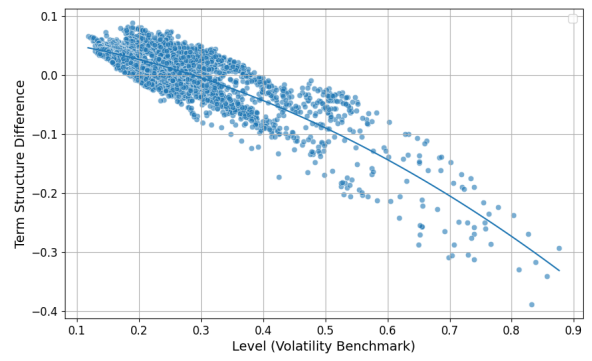
(e) 5M



(f) 6M

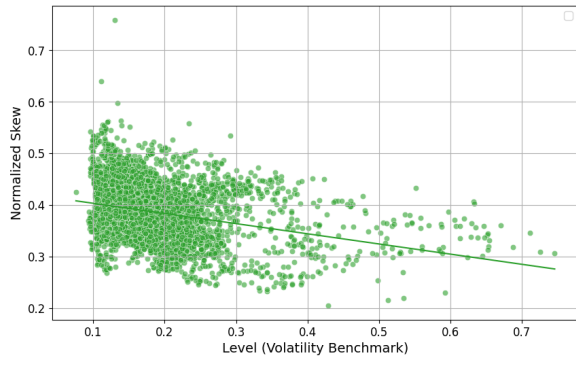


(g) 9M

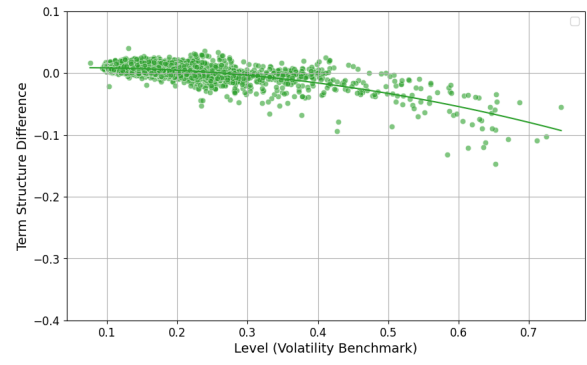


(h) 12M

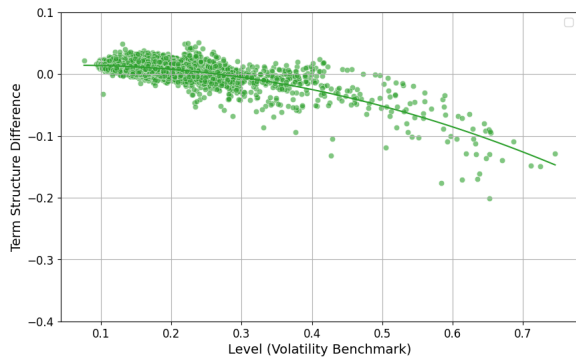
Figure 13: Normalized skew and term structure regressions on the level of the volatility benchmark for the Russell 2000.



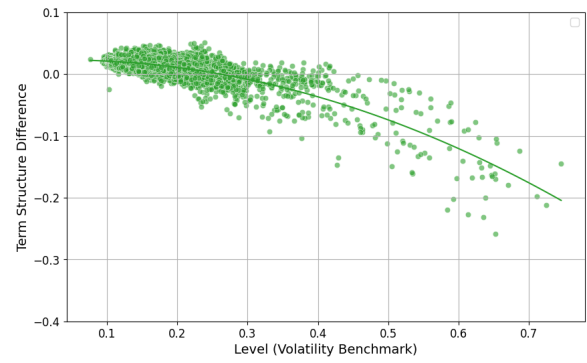
(a) Normalized skew



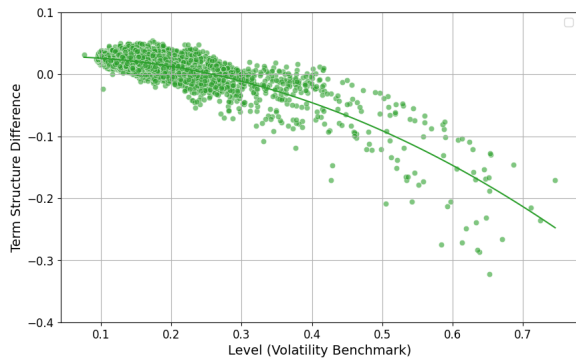
(b) 2M



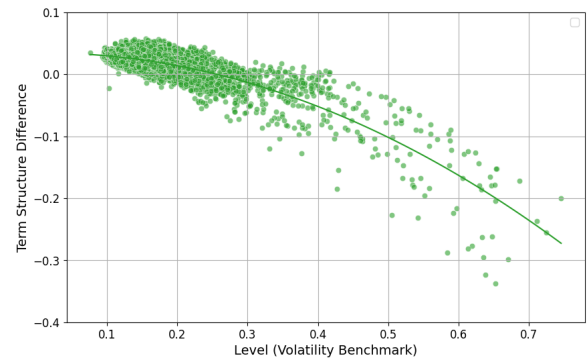
(c) 3M



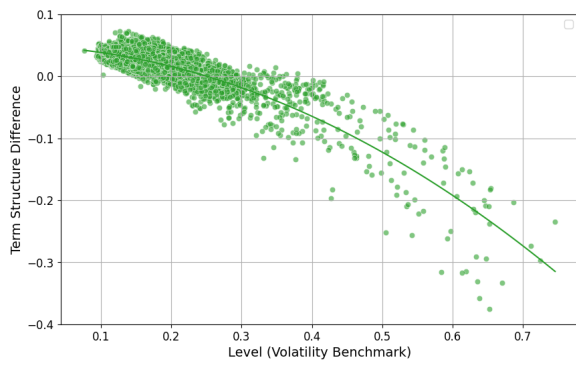
(d) 4M



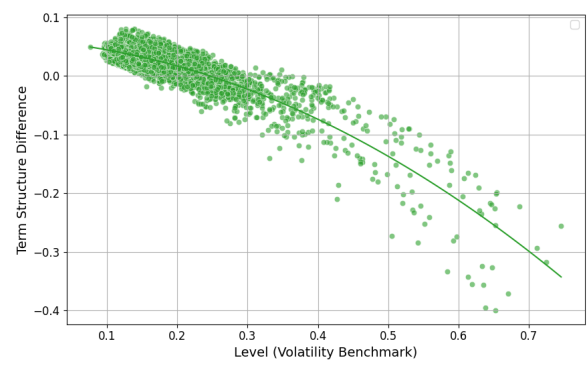
(e) 5M



(f) 6M

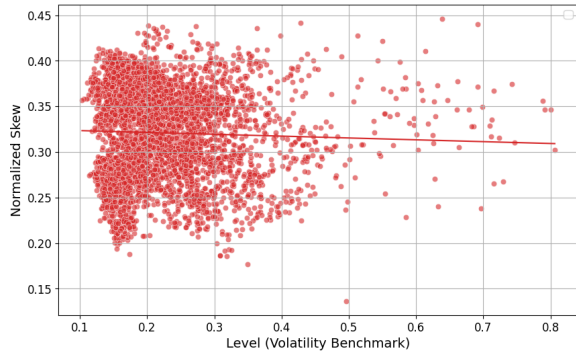


(g) 9M

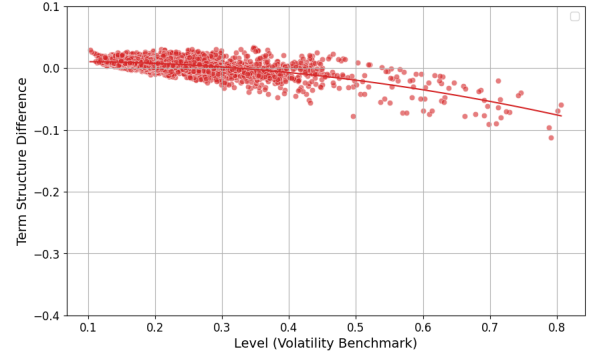


(h) 12M

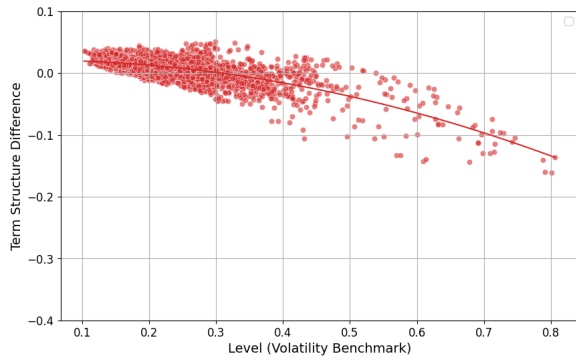
Figure 14: Normalized skew and term structure regressions on the level of the volatility benchmark for the Dow Jones Industrial Average.



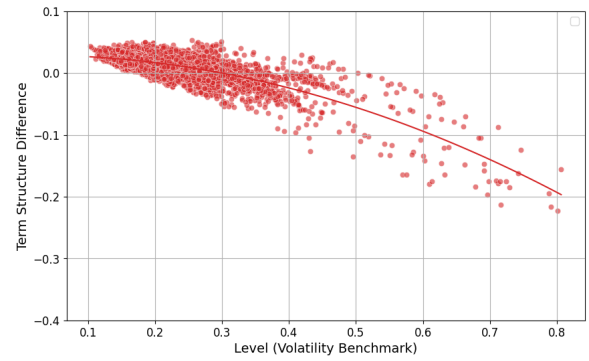
(a) Normalized skew



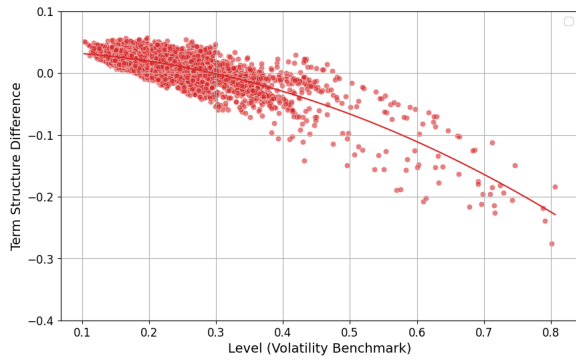
(b) 2M



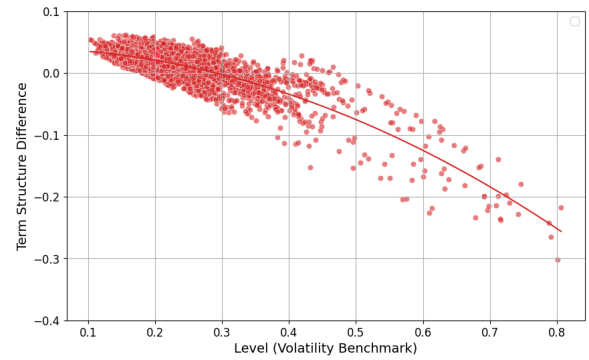
(c) 3M



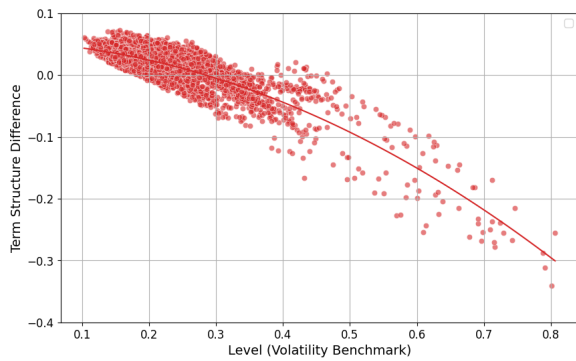
(d) 4M



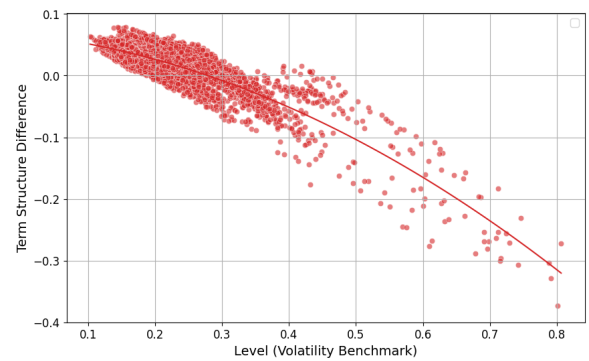
(e) 5M



(f) 6M



(g) 9M



(h) 12M

Figure 15: Normalized skew and term structure regressions on the level of the volatility benchmark for the Nasdaq-100.

References

- Alm, T., Harrach, B., Harrach, D. & Keller, M. (2013). A monte carlo pricing algorithm for autocallables that allows for stable differentiation. *Journal of Computational Finance*, 17(1), 43–70.
- Ametrano, F. & Ballabio, L. (2020). *Quantlib: A free/open-source library for quantitative finance*. (Retrieved from <https://www.quantlib.org>)
- Anderson-Cook, C. M. (2006). *Quantitative risk management: concepts, techniques, and tools*. Taylor & Francis.
- Ang, C. S. (2015). *Analyzing financial data and implementing financial models using r*. Springer.
- Bauer, R. (2012). *Fast calibration in the heston model* (Unpublished doctoral dissertation). Technische Universität Wien.
- Bergomi, L. (2015). *Stochastic volatility modeling*. CRC press.
- Boyle, P., Broadie, M. & Glasserman, P. (1997). Monte carlo methods for security pricing. *Journal of Economic Dynamics and Control*, 21(8-9), 1267–1321.
- Cox, J. C., Ingersoll Jr, J. E. & Ross, S. A. (1985). An intertemporal general equilibrium model of asset prices. *Econometrica: Journal of the Econometric Society*, 363–384.
- Crisóstomo, R. (2015). An analysis of the heston stochastic volatility model: Implementation and calibration using matlab. *arXiv preprint arXiv:1502.02963*.
- De Col, A. & Kuppinger, P. (2017). The interplay between stochastic volatility and correlations in equity autocallables. *Available at SSRN 3228065*.
- DeMiguel, V., Garlappi, L. & Uppal, R. (2009). Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *The Review of Financial Studies*, 22(5), 1915–1953.
- Deng, G., Mallett, J. & McCann, C. (2011). Modeling autocallable structured products. *Journal of Derivatives & Hedge Funds*, 17, 326–340.
- Dimitroff, G., Lorenz, S. & Szimayer, A. (2011). A parsimonious multi-asset heston model: Calibration and derivative pricing. *International Journal of Theoretical and Applied Finance*, 14(08), 1299–1333.
- Dupire, B. (1994). Pricing with a smile. *Risk*, 7(1), 18–20.

- Farkas, W., Ferrari, F. & Ulrych, U. (2024). Pricing autocallables under local-stochastic volatility. In *Peter carr gedenkschrift: Research advances in mathematical finance* (pp. 329–378). World Scientific.
- Ge, W. (2016). A survey of three derivative-based methods to harvest the volatility premium in equity markets. *The Journal of Investing*, 25(3), 48–58.
- Glasserman, P. & Staum, J. (2001). Conditioning on one-step survival for barrier option simulations. *Operations Research*, 49(6), 923–937.
- Guillaume, T. (2015). Autocallable structured products. *The Journal of Derivatives*, 22(3), 73–94.
- Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies*, 6(2), 327–343.
- Hsieh, M.-H., Liao, W.-C. & Chen, C.-L. (2014). A fast monte carlo algorithm for estimating value at risk and expected shortfall. *Journal of Derivatives*, 22(2), 50.
- Ilmanen, A. (2012). Do financial markets reward buying or selling insurance and lottery tickets? *Financial Analysts Journal*, 68(5), 26–36.
- Israelov, R., Nielsen, L. N. & Villalon, D. (2017). Embracing downside risk. *The Journal of Alternative Investments*, 19(3), 59.
- Jäckel, P. (2002). *Monte carlo methods in finance*. West Sussex: John Wiley & Sons, Ltd.
- Kahneman, D. & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47, 263–291.
- Kim, J. & Finger, C. C. (2000). A stress test to incorporate correlation breakdown. *Journal of Risk*.
- Kirby, C. & Ostdiek, B. (2012). It’s all in the timing: simple active portfolio strategies that outperform naive diversification. *Journal of Financial and Quantitative Analysis*, 47(2), 437–467.
- Ledoit, O. & Wolf, M. (2008). Robust performance hypothesis testing with the sharpe ratio. *Journal of Empirical Finance*, 15(5), 850–859.
- Lo, A. W. (2002). The statistics of sharpe ratios. *Financial Analysts Journal*, 58(4), 36–52.

- Lord, R., Koekkoek, R. & Dijk, D. V. (2010). A comparison of biased simulation schemes for stochastic volatility models. *Quantitative Finance*, 10(2), 177–194.
- Mixon, S. (2011). What does implied volatility skew measure? *Journal of Derivatives*, 18(4), 9.
- Politis, D. N. & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428), 1303–1313.
- Salon, G. (2019). Equity autocalls and vanna negative carries: pricing and hedging with a simple add-on. *Available at SSRN 3383122*.
- Sobol, I. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7, 86–112.
- Szymanowska, M., Horst, J. T. & Veld, C. (2009). Reverse convertible bonds analyzed. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 29(10), 895–919.
- Tzeng, Y.-Y., Beaumont, P. M. & Ökten, G. (2018). Time series simulation with randomized quasi-monte carlo methods: An application to value at risk and expected shortfall. *Computational Economics*, 52, 55–77.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D. et al. (2020). Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature Methods*, 17(3), 261–272.
- Wallmeier, M. & Diethelm, M. (2009). Market pricing of exotic structured products: The case of multi-asset barrier reverse convertibles in switzerland. *Journal of Derivatives*, 17(2), 59.
- Wilkins, S., Erner, C. & Röder, K. (2003). The pricing of structured products—an empirical investigation of the german market. *The Journal of Derivatives*, 11(1), 55–69.