UNIVERSITY OF SUSSEX

BUSINESS SCHOOL

Trading Bitcoin Options

*Thesis submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy*

ARBEN IMERAJ

12 JUNE 2023

Abstract

This thesis provides a fundamental analysis of the bitcoin implied volatility. The bitcoin options market is highly fragmented, split between regulated and unregulated exchanges. This variety of different exchange types creates plain vanilla options with various settlement procedures. In this dissertation, we present standard or direct, inverse and quanto inverse options, explain their pricing structure, and conduct a delta-hedging study considering various smile-dependent and smile-adjusted delta-hedge ratios. All research is conducted on a unique dataset of option prices which we acquire directly from the exchange. It contains all option, futures, and perpetuals trade data over the past three years on an intra-day frequency; and, with over ten million option prices, it is unmatched by any other research at the time of writing. We introduce an implied volatility index, the first of its kind for crypto, and consider its role as a fair value variance swap rate for different maturities, reflecting the distinct trading behaviour and clustering of bitcoin options. Moreover, we provide insight of the determinants of the bitcoin implied volatility and compare it to frequently traded volatility indices of other asset classes. Standard mechanisms such as margin calls do not exist in unregulated crypto derivatives markets, impacting the implied volatility of this emerging asset class directly. We take a detailed look at the bitcoin implied volatility surface, highlight key characteristics, and point out differences to other asset classes such as equity, commodity, or currencies. We capture the surface dynamics through a second-order Taylor expansion, model the resulting time series of parameters with statistical and machine learning models, both of which are widely used in the industry, and derive profitable trading strategies. In many aspects, this work is ahead of its time, but our findings help practitioners understand the market better and price derivatives correctly. We pave the way for further research and allow other academics to build on our analyses.

Keywords: Cryptocurrency, Derivatives, Futures, Implied Volatility, Options, Realised Volatility, Fear Gauge, Inverse Option, Foreign Exchange, Hedging, Derivatives Hedging, Perpetual Contract, Robust Finance, Dynamic Delta Hedging

Declaration of Original Authorship

I hereby declare that this thesis is my own work and that no other sources have been used except those clearly indicated and referenced. Where research has been carried out in collaboration with, or supported by others, this is duly acknowledged below, and my contribution is indicated. This thesis has not been and will not be, submitted to another university for the award of any other degree.

Arben Imeraj

12 June 2023

Details of Collaboration and Other Sources:

- Chapter 4, particularly the conceptualisation of Subsection 4.3.3, is co-authored with Prof. Carol Alexander and Dr. Ding Chen. All other tasks related to derivations, preparation, and analyses are entirely my own work. The write-up and presentation of results have been carried out collectively with my co-authors. This work has been peer-reviewed and is accepted for publication (Alexander et al., 2023).

- Chapter 5 is co-authored with Prof. Carol Alexander. All tasks related to data gathering, preparation, and statistical analyses are entirely my own work. The write-up and presentation of results have been carried out collectively with my co-author. This work has been peer-reviewed and published (Alexander and Imeraj, 2023).

- Chapter 6 is co-authored with Prof. Carol Alexander. All tasks related to data gathering, preparation, and statistical analyses are entirely my own work. The write-up and presentation of results have been carried out collectively with my co-author. This work has been peer-reviewed and published (Alexander and Imeraj, 2021).

- This dissertation incorporates various components of my publications mentioned above, such as text segments, definitions, figures, tables or formulae. These elements have been adapted in full, updated to contemporary data or paraphrased also in other chapters of this thesis. In an effort to maintain fluency and readability, comprehensive acknowledgements for each individual element extracted from my publications are consciously omitted.

- Chapter B of the Appendix is a textbook chapter. All ideas, definition, formulae and some figures are inspired by or adopted from Goodfellow et al. (2016), Kelleher (2019), Schmidt (2019) Jansen (2020), Brownlee (2021) and Zhang et al. (2023). In an effort to maintain fluency and readability, comprehensive acknowledgements for each individual element extracted from the papers and books are consciously omitted.

ACKNOWLEDGEMENTS

First and foremost, I am deeply grateful to my supervisors Carol Alexander and Ding Chen for their constant guidance, enduring support, and continuous tutelage over the last three and a half years. Truthfully, it has been both a pleasure and honour to work alongside such experienced researchers and brilliant people which has helped me grow tremendously in many ways. I am convinced that their influence will be a huge aid to succeed in my future endeavours.

I am indebted to my dear friend and fellow PhD student Daniel Heck, who I shared not just a roof over the past years, but the highs and lows of my PhD journey. His advice and camaraderie has provided me with invaluable support and made both the triumphs sweeter and challenges lighter.

My heartfelt thanks to Flaka Shoshi, whose passion for knowledge inspired my own PhD journey. Her achievements in academia lit the way, showing me the heights that dedication and curiosity can reach. Her example has been a constant source of motivation, driving me to aim higher.

Finally, I would like to thank those closest to my heart: my family. For it is their unconditional love and support throughout my entire life that inspired me and brought me where I am today. Dedicating this thesis to them is merely a humble gesture towards a debt I will never be able to fully repay.

Contents

List of Figures

List of Tables

ACRONYMS

ADF   Augmented Dickey-Fuller.
AMM   Automated Market Maker.
API   Application Programming Interface.
ATM   at-the-money.

bps   basis points.
BRR   CME CF Bitcoin Reference Rate.
BS   Black-Scholes Model.
BTC   bitcoin.

CBOE   Chicago Board Options Exchange.
CFTC   Commodities and Futures Trading Commission.
CME   Chicago Mercantile Exchange.

DBTC   Deribit Bitcoin Index.
DeFi   Decentralised Finance.
DEX   Decentralised Exchanges.
DOV   DeFi Option Vaults.
DTE   Days to Expiration.
DVOL   Deribit Implied Volatility Index.

EMM   Equivalent Martingale Measure.
ETH   ether.

FX   Foreign Exchange.

GBM   Geometric Brownian Motion.

ITM   in-the-money.
IV   implied volatility.
IVS   implied volatility surface.

LP	Liquidity Pool.
LSTM	Long-Short Term Memory.
MAE	Mean Absolute Error.
MCD	Mean Correct Direction.
MCDO	Monte Carlo Dropout.
MiCA	Markets in Cryptoassets.
ML	Machine Learning.
MV	Minimum Variance.
NN	Neural Networks.
OLS	Ordinary Least Square.
OTC	over-the-counter.
OTM	out-of-the-money.
PCA	Principal Component Analysis.
PCHIP	Piecewise Cubic Hermite Interpolation Polynomials.
PCP	Put-Call parity.
PnL	Profit and Loss.
RMSE	Root Mean Squared Error.
RV	Realised Volatility.
RW	Random Walk.
SEC	US Securities and Exchange Commission.
SM	Sticky Money.
SS	Sticky Strike.
ST	Sticky Tree.
SV	stochastic volatility.
SVI	stochastic volatility inspired.
TTM	Time to Maturity.
USD	US Dollar.
USDC	USD Circle Coin.
USDT	tether.
UST	TerraUSD.
VAR	Vector Auto-Regressive.
VRP	Variance Risk Premium.

Abstract

This thesis provides a fundamental analysis of the bitcoin implied volatility. The bitcoin options market is highly fragmented, split between regulated and unregulated exchanges. This variety of different exchange types creates plain vanilla options with various settlement procedures. In this dissertation, we present standard or direct, inverse and quanto inverse options, explain their pricing structure, and conduct a delta-hedging study considering various smile-dependent and smile-adjusted delta-hedge ratios. All research is conducted on a unique dataset of option prices which we acquire directly from the exchange. It contains all option, futures, and perpetuals trade data over the past three years on an intra-day frequency; and, with over ten million option prices, it is unmatched by any other research at the time of writing. We introduce an implied volatility index, the first of its kind for crypto, and consider its role as a fair value variance swap rate for different maturities, reflecting the distinct trading behaviour and clustering of bitcoin options. Moreover, we provide insight of the determinants of the bitcoin implied volatility and compare it to frequently traded volatility indices of other asset classes. Standard mechanisms such as margin calls do not exist in unregulated crypto derivatives markets, impacting the implied volatility of this emerging asset class directly. We take a detailed look at the bitcoin implied volatility surface, highlight key characteristics, and point out differences to other asset classes such as equity, commodity, or currencies. We capture the surface dynamics through a second-order Taylor expansion, model the resulting time series of parameters with statistical and machine learning models, both of which are widely used in the industry, and derive profitable trading strategies. In many aspects, this work is ahead of its time, but our findings help practitioners understand the market better and price derivatives correctly. We pave the way for further research and allow other academics to build on our analyses.

CHAPTER 1

INTRODUCTION

Blockchain technology, on which every cryptocurrency is built, is renaître. The storage of information on a distributed ledger which is visible to all participants goes back to the 1990s at least (Haber and Stornetta, 1991, 1997). But the idea of a cryptographically-linked blockchain technology only grew in popularity after the introduction of a cryptocurrency called bitcoin (BTC) as the native token for a blockchain called Bitcoin (Nakamoto, 2009). Bitcoin transactions are verified by a peer-to-peer network, stored on the blockchain, transparent to anyone with internet access, and become immutable after a certain period of time. Today, blockchains of many different types have exceeded their initial purpose of recording simple peer-to-peer transactions, becoming the backbone of Web 3.0 by carrying fully autonomous and self-regulating smart contracts. As a result, there are tens of thousands of tokens which are transacted on blockchains and also off-chain on market places and centralised exchanges.

Although the initial idea of bitcoin was to serve as a decentralised alternative to central bank issued fiat currencies and to act as a medium of exchange, today's cryptocurrency markets have virtually nothing to do with this concept. Instead, the variety of blockchain architectures has evolved cryptocurrencies into a tradable asset class of their own. Initially a paradise for speculators, facing a highly volatile and inefficient market (Urquhart, 2016), populated by retail traders hoping to buy low and sell high, the crypto market has matured over the past years to a highly liquid and efficient market hosting both retail trader and (professional) institutional investors.[1] Despite the ongoing debates surrounding cryptocurrencies' intrinsic value, their contributions for society or their immense energy consumption, trading cryptoassets – particularly derivatives – and institutional involvement are at record levels, proving a wide-scale acceptance of this asset class.

---

[1]BlackRock, Fidelity and Charles Schwab started offering their clients investment opportunities in bitcoin in form of simple money markets or exchange-traded funds or notes. Top tier banks like Goldman Sachs or J.P. Morgan and proprietary trading houses such as Jump and Cumberland Capital have been active in crypto spot and derivatives market for many years. See Financial Times accessed on 01 May 2023.

Hence, it is somewhat surprising that regulators around the world continue to lack a coherent framework or strategic plan for regulating this multi-trillion dollar market.[2] Instead, there is still an ongoing feud between the US Securities and Exchange Commission (SEC) and the Commodities and Futures Trading Commission (CFTC) concerning regulatory jurisdiction over cryptocurrencies, which revolves around the classification of tokens – specifically, which should be considered securities and which commodities. At the same time, the European Union adopted its own comprehensive set of rules for cryptoassets, i.e. the Markets in Cryptoassets (MiCA) regulation, set to come into force between 2024 and 2025. However, the slow progress of this law and the 'wait-and-see' approach of the Europeans, hoping to get their transatlantic partners on board, drives both private and institutional investors towards unaudited exchanges, so-called 'unregulated' exchanges, with high risk of criminal activity. This lack of regulation has not only a far-reaching impact on both retail and institutional traders, but poses systemic risk for the general financial system as the collapse of the FTX exchange in November 2022 has demonstrated.[3]

These unregulated exchanges, or 'self-regulated' as they often label themselves, typically operate outside the bounds of governmental oversight and prefer running their operational business within jurisdictions that offer a financial safe harbour like the Bahamas or Panama. Typically, these exchanges are centralised in a sense that they facilitate trades through an electronic limit order book system.[4] They operate as a shadow bank, custodian for client funds, brokerage firm, and clearing and settlement house, all in one, and allow (retail) investors to trade spot with up to tenfold leverage. Most prominent exchanges in this class include Binance, OKX or Huobi. On the other side of the regulatory spectrum, exchanges like Coinbase, LMAX or Kraken comply to, and cooperate with the financial authorities of the respective jurisdiction they operate in. These (semi-)regulated exchanges facilitate the same centralised trading but open their doors and books to auditors and governmental representatives. However, according to the latest CCData Report, the overall traded spot volume in March 2023 was just short of 900 billion US Dollar (USD), and over 80% of these trades were facilitated by unregulated exchanges, underscoring their dominance in this financial arena.

The same lack of regulation applies to cryptocurrency derivatives trading, which is why this market is often called the 'wild west'. Today's crypto derivatives markets are highly fragmented and split between regulated and unregulated exchanges. Similar to spot trading, the vast majority of derivatives exchanges offer electronic limit order book trading on a variety

---

[2]As of May 2023, the accumulated market capitalisation of all cryptocurrencies is well above $1.2 trillion, see Statista.com.

[3]Plenty of traditional asset manager, pension funds, VC companies or law firms got caught in the aftermath of the FTX collapse loosing billions of US dollar, see Financial Times. The Bank of England deputy governor points out the risk associated with crypto for the financial system, see this article, accessed on 01 May 2023.

[4]Decentralised Exchanges (DEX) provide an alternative type of exchange to facilitate token swaps. These are blockchain protocols, or smart contracts, which allow participants to deposit their token pair, e.g. ETH and USDC, into Liquidity Pool (LP) in return for a premium or reward. An Automated Market Maker (AMM) provides a relative price of one token with respect to the other, and facilitate fully autonomous trades based on the size of the LPs. Popular protocols include UniSwap or Curve.

of underlying assets and different contract sizes.[5] The Chicago Board Options Exchange (CBOE) was the first US exchange to introduce bitcoin futures in December 2017 but stopped issuing new maturities barely half a year later due to lack of interest. At the same time, the Chicago Mercantile Exchange (CME) introduced cash-settled futures on the CME CF Bitcoin Reference Rate (BRR), a volume-weighted reference rate representing bitcoin's dollar price across the most prominent spot exchanges. Since then, both volume and open interest have been steadily increasing, especially following the launch of the micro bitcoin futures in May 2021, which have a notional value equal to a tenth of a bitcoin. This is in stark contrast to the standard futures with a notional value of five bitcoins, aiming to attract more participants. However, regulated exchanges fail to attract the main players in this sector. In March 2023, the CME reported a traded volume of 44.8 billion USD and an open interest of 1.74 billion USD. In contrast, unregulated exchanges recorded a notional volume of 1.3 trillion USD with an aggregate open interest of over nine billion USD. Notably, Binance alone accounted for over 758 billion USD traded futures notional in March 2023.

One reason for this excess trading on self-regulated exchanges lies in their unique characteristics, such as the opportunity to trade 24/7, every day of the year, and a notional amount as low as a single dollar, which can be particularly appealing to retail traders. Furthermore, they provide an extended spectrum of tradable products, most notably the so-called 'perpetual contracts', or perpetuals for short, which is a unique product tradable only on crypto derivatives markets. These derivatives mimic a margin-based spot account, allowing trading of the underlying spot with up to 125 times leverage. Although similar in structure to standard futures, perpetuals distinguish themselves by the absence of an expiry date. Yet, their funding mechanism prevents these products from deviating too far from the underlying. This mechanism involves a regular reset of perpetual prices which occurs every eight hours, wherein long and short positions exchange cashflows based on the difference between the perpetual and spot price. This has led to these products also being referred to as 'perpetual swaps'. Perpetual contracts are by far the most popular products in the realm of cryptocurrencies, likely due to the lack of roll-over risk and low marginal requirements. Indeed, these products open a new research strain as they can be used for hedging, alternative to the underlying or fixed-maturity futures as Alexander et al. (2020) and Alexander et al. (2023) show.

A very distinct feature of unregulated exchanges is that they never introduced the concept of margin calls. Instead, any leveraged position in futures or perpetuals that trends adversely, drying up the margin account, is automatically and immediately liquidated. This involves that the position is closed at prevailing market prices, without the possibility to deposit

---

[5]At the other end of the spectrum, Decentralised Finance (DeFi) protocols like dYdY or Ribbon Finance provide the ground to trade futures and option strategies without an intermediary. The dYdX exchange provides a decentralised alternative platform for futures trading and mimics a limit orderbook though LPs. On the options side, Ribbon Finance, alongside Thetanuts, offer traders to invest in DeFi Option Vaults (DOV) which are smart contracts mimicking option strategies such as covered calls or cash-covered puts. Traders deposit/stake collateral into vaults and market maker try to buy vault tokens and pay a premium which is distributed among the investors. However, with $\sim \$160m$ total volume in March 2023 this market is too small to consider, see DeFi Llama.

more liquidity into the margin account. This forced selling can potentially result in a vicious cycle of price depreciation, leading to further liquidations and potentially contributing to bitcoin's infamous and distinct volatility. Modern cryptocurrency markets have the tools and means to trade this distinct feature of bitcoin directly through options and, since recently, via futures on implied volatility indices. Options on bitcoin and other cryptocurrencies have been available since 2017, but it was not until recently that their trading volumes and interest took a substantial leap. With a traded volume exceeding 35 billion USD in notional terms in March 2023, it has become a market too big to ignore. Some major financial institutions even forecast that bitcoin options will constitute "the next big step" in this domain.[6]

The exchange fragmentation has led bitcoin options to become a tale of two settlement mechanisms, i.e. standard – or direct – and inverse. The former describe plain vanilla European options on the bitcoin price, an index value or some reference rate, e.g. BRR. The distinction between standard and direct depends on whether options are denominated using fiat currencies (standard) or those paired with a stablecoin, such as tether (USDT) which we denote as direct.[7] The CME launched its first European cash-settled option on their bitcoin futures in January 2020. However, the contract size of five bitcoin proved to be a deterrent, and the anticipated influx of institutional investors did not occur. This is why, in March 2022, the CME launched micro bitcoin options with a contract size of a tenth of a bitcoin, aiming to compete with the self-regulated platforms and target retail traders. Despite the CME introducing a wide spectrum of reference rates for other cryptocurrencies – such as polkadot (DOT) or solana (SOL) – they currently offer only bitcoin and ether (ETH) options. Self-regulated exchanges like Binance or OKX offer direct European cash-settled options for bitcoin and ether, in addition to the Binance native token (BNB) and Ripple (XRP). Contrary to standard options, the underlying for these direct options is denominated in a stablecoin, primarily USDT and USD Circle Coin (USDC), and hence avoid any fiat transaction.

On the other hand, exchanges like BitMEX and Deribit have introduced 'inverse' products, which include futures, perpetuals and options that are margined and settled in bitcoin despite the underlying being the USD or USDT value of bitcoin or any proxy. That is, the payoff and premium of a bitcoin option are paid in bitcoin, but the tracked underlying is denominated in another currency. To obtain the maturity payoff, one needs to calculate the difference between the bitcoin value in dollars and the option strike (also quoted in dollars), then convert the result to bitcoin using the bitcoin value at maturity. By adopting this approach, Deribit and similar exchanges have facilitated a fully functional options market without the necessity of onboarding any fiat currency, thus circumventing potential regulatory or legal complications. Evidently, this is a highly demanded feature as the trading volumes and proportions indicate. While the CME reported a trading volume exceeding 1.6 billion USD in March 2023, unregulated exchanges recorded a volume surpassing 35 billion USD, with Deribit accounting for

---

[6]See Bitcoin Options Volume and Goldman Sachs Bitcoin Options accessed 01 May 2023.

[7]A stablecoin describes a token which is pegged one-to-one to a real asset like fiat currencies or gold. However, in this context, we consider stablecoins as USD-pegged tokens and ignore other currencies/assets.

over 88% of this total. In terms of open interest, Deribit indisputably dominates the market possessing over 85% of the total market share. At present, inverse options are available for bitcoin, ether, and solana, but given the trends over recent years, it is only a matter of time that options for additional underlying coins will soon be offered.

Both settlement mechanisms for crypto options, which differ between exchanges, as well as the different types of products, are not yet widely understood and a clear mathematical exposition of these products is lacking. Given the increasing acceptance and growing interest in cryptocurrency options, it is important to fully understand the pricing mechanisms of bitcoin option types. In Chapter 4, we close the gap in financial literature and provide a sound explanation of the different settlement mechanisms. We highlight key differences between standard or direct and inverse options, derive the correct pricing formula, and show that inverse options can and should be priced in accordance to the pricing model for Foreign Exchange (FX) rates by Garman and Kohlhagen (1983). Motivated by the instability of stablecoins (Duan and Urquhart, 2023), we draw attention to the risks of direct options which use stablecoins rather than fiat currencies for settlement and propose quanto options to mitigate the risk of de-pegging. Lastly, we introduce a novel exotic option mimicking the inverse payoff and providing an attractive, less expensive alternative for retail traders eager to participate in the bitcoin options market.

As established industry giants such as Goldman Sachs or BlackRock have commenced collaboration with cryptocurrency exchanges and introduced digital asset products to the wider public, the necessity for hedging these assets has become increasingly essential.[8] This could potentially account for the fact that the CME has started to catch up on unregulated exchange volumes. For instance, the proportion of traded notional options volume (open interest) on the CME in March 2022 was approximately 3% (6%), whereas by March 2023, it had increased to 6% (10%), suggesting further a growing trend. While banks and asset managers may resort to futures and perpetuals to hedge their structured products, options offer a versatile alternative through delta-hedging. Typically, a delta-hedged position mitigates exposure to changes in the underlying, with the Black-Scholes Model (BS) model offering a fundamental understanding of option price modelling and hedging. However, numerous empirical studies in financial literature have contested the validity of the assumptions underpinning the BS model. Unintentionally, bitcoin option markets come closest to satisfying the BS model assumptions, given their 24/7 operation, thereby – theoretically – allowing to hedge option position continuously. Nevertheless, constant volatility in both traditional and crypto markets is not given, with implications on the delta-hedging position. Traders must take into account the correlation between the underlying and its implied volatility (IV) to execute effective hedging. Given the distinct nature of bitcoin volatility, a study on dynamic delta-hedging would yield invaluable insights for traders and further expand the understanding of this emerging market. Yet, thus far, a mere handful of researchers (Teng and Härdle, 2022; Matic et al., 2023) have actively

---

[8]See Bloomberg accessed on 01 May 2023.

explored BTC option hedging, and none looked into robust and efficient delta adjustments.

Chapter 5 will shed light on the dynamical delta-hedging behaviour of bitcoin and provide fundamental knowledge of the distinct behaviour of the implied volatility. We compare two different hedging instruments, the standard fixed-maturity futures and the novel perpetual swaps, and evaluate their performance over the longest available dataset. Unlike other research, the deltas considered here are smile-adjusted and smile-implied, and minimise the instantaneous variance of a delta-hedged portfolio. Additionally, this study does not limit itself to daily rebalancing but instead implements intra-day position adjustments. The chapter delivers two significant contributions. Firstly, it highlights the regime-dependency of bitcoin's implied volatility and refutes the leverage effect for cryptocurrency markets, thus influencing the performance of individual smile-adjustment models. Secondly, although we do not identify a single model that consistently surpasses the BS delta across all strike levels based on the variance of hedging errors, our findings suggest promising results for at-the-money (ATM) and out-of-the-money (OTM) tails and incentivise market makers to use the proposed models.

A delta-hedged option not only mitigates exposure to the underlying, but it also represents the simplest form of volatility trading. The increasing volume in traded contracts can likely be attributed to both institutional hedgers and volatility traders. While traditional markets provide alternatives for direct volatility trading, such as implied volatility indices or variance swaps, these products are only beginning to make their appearance in the crypto market. Just recently, the Deribit exchanged launched futures on a bitcoin implied volatility index, granting traders direct access to volatility trading. Prior to that, the proprietary trading house GSR launched the first ever bitcoin variance swap in 2020, which is yet another indicator for the growing interest and the prosperous future of the crypto options market. Both variance swaps and implied volatility indices are mathematically related as the methodology of modern volatility indices, e.g. VIX, is build upon a fair valuation of variance swaps. Consequently, the index can be interpreted as a fair variance swap rate, and the payoff to that swap can be defined as Variance Risk Premium (VRP), see for example Bakshi and Kapadia (2003). Such an index summarises the implied volatility skew in a single number and indicates sophisticated option trader's expectation of future volatility. While there exists a plethora of empirical financial literature focusing on the variance risk premium of US equity indices (Carr and Wu, 2009), this type of analysis remains rather limited for cryptocurrency markets.

This motivates us to introduce a bitcoin implied volatility index, the first of its kind, using a variance swap based methodology similar to the CBOE VIX. In Chapter 6, we construct this index for various maturities, taking into account distinct trading clusters of bitcoin options, and use this index as a variance swap rate to examine bitcoin's variance risk premium. Furthermore, we test a different index methodology put forth by Leontsinis and Alexander (2017), which aims to reduce the drawbacks of the variance swap methodology and compare the variance risk premiums derived from this methodology with the traditional CBOE derivation. In an attempt to better understand bitcoin's role in the global economy, we compare

our results to variance risk premia of other asset classes. Our findings bear significance and contribute to the existing literature in three ways. First, they reveal that bitcoin displays VRP patterns similar to other asset classes, i.e. typically negative but with abrupt surges into positive territory. Second, we show that bitcoin's implied volatility tends to be overvalued and proves to be an unreliable predictor of realised volatility. Lastly, bitcoin has, over long periods in the past, surprisingly exhibited the most attractive variance risk premium profile across all compared asset classes, thus representing a great opportunity for swap writers.

Having discussed the implied volatility behaviour of bitcoin in great detail, and leveraged the information content of the implied skew, we aim to forecast the bitcoin implied volatility surface (IVS). The capacity to anticipate and control the dynamics of future shifts in implied volatility is critically important for pricing and hedging options, as well as for trading systematic volatility. This strain of research has sparked substantial interest among both practitioners and scholars (Gonçalves and Guidolin, 2006). Similar to previous discussions, the predictability of equity IVS and other asset classes have been studied widely (Cont and Da Fonseca, 2002; Bernales and Guidolin, 2014), but only little is known about their effectiveness on bitcoin. Chapter 7, the final research chapter, focuses on forecasting the bitcoin implied volatility surface and leveraging it for trading purposes. In fact, this involves a two-step approach in which we first model the surface, and subsequently predict the model parameter. Reflecting the increasing interest in machine learning techniques in recent literature, we incorporate a recurrent neural network and compare its performance against a statistical time series model and a naive approach. Through this chapter, we bridge the existing gap in literature and test the market effectiveness of bitcoin options. Despite our approach being robust and easily implementable, our findings show that the surface is predictable and expose potential profitable trading strategies based on our model's forecast. Nevertheless, real market factors such as transaction costs and the bid-ask spread negate these profits, pointing towards an efficient bitcoin options market.

The remaining Chapter 2 provides a comprehensive literature review, focusing primarily on the body of research that explores options and volatility trading in the context of cryptocurrencies. This section also describes the theoretical models employed in the study. Chapter 3 explains the processing procedure of our proprietary bitcoin options market data and describes the data filtering and interpolation techniques used. Furthermore, this chapter presents an analysis of options trading statistics spanning a three-year period, which represents the most extensive dataset currently available on bitcoin options trading. The final chapter of this dissertation concludes by summarising the primary results and provides a prospectus for potential future research in the field. The appendices include a derivations and calculations, a detailed exposition of machine learning models with a focus on neural networks as drawn from relevant textbook resources, and a review of the robustness checks employed. Across this thesis, we will introduce a different `font` to refer to used `Python` libraries.

CHAPTER 2

LITERATURE REVIEW

Prior to engaging in any research on bitcoin options pricing, hedging or bitcoin implied volatility behaviour, it is essential to determine the appropriate pricing mechanisms of these derivatives and gain an understanding of the diverse settlement mechanisms. There exists a wide body of financial literature examining inverse futures contracts on bitcoin and other cryptocurrencies in great detail (Deng et al., 2020, 2021). Notably, Alexander et al. (2020) (perpetuals) and Alexander et al. (2023) (fixed-maturity) provide comprehensive insights into the various types of futures. The authors outline key distinctions between inverse contracts and their traditional counterparts, explain the inverse settlement mechanism in-depth, and present the pricing methods before revealing their findings on the hedging behaviour of inverse futures. While cryptocurrency options too can be of inverse or direct type, the majority of existing academic research neglects to differentiate between these settlement types. Even fewer studies address this distinction actively. Although direct options are no innovative product and can be modelled using any 'traditional' deterministic or stochastic volatility (SV) options model such as Black and Scholes (1973) or Heston (1993), inverse options necessitate a more sophisticated approach. In fact, at the time of writing, only Lucic (2022) and Sepp and Rakhmonov (2022) offer a concise discussion on the key characteristics of inverse options.

2.1 Pricing

In his two-page working paper, Lucic (2022) indicates that inverse options can be priced within a standard FX framework. He assumes the BTCUSD rate follows a Geometric Brownian Motion (GBM) and that functioning money markets for both bitcoin and USD exist, with corresponding martingale measures (domestic and foreign). Following standard risk-neutral evaluation arguments (Musiela and Rutkowski, 2005), the author links these two martingale measures and presents a relationship between inverse and direct options, akin to the duality

condition observed in FX options. Despite accurate pricing, Lucic (2022) fails to provide a motivation for adopting this approach, omits an explanation of the practical implications for traders using this pricing mechanism, and disregards the incomplete market on the Deribit exchange. The author's methodology lacks specification of interest rates or information on the volatility of the underlying. However, it is plausible to assume constant rates and volatility in line with the assumptions proposed by Black and Scholes (1973). In essence, the content of this paper suggests that pricing inverse options resembles the FX methodology put forth by Garman and Kohlhagen (1983) closely, even though this is not explicitly stated.

This pricing approach can be readily extended beyond the Black-Scholes world. Solutions to option pricing problems are attainable, at least in the Fourier transform sense, for any tractable Lévy processes governing the evolution of the BTCUSD exchange rate. In this context, Sepp and Rakhmonov (2022) extend the work of Lucic (2022), offering a more comprehensive description of option valuation under simple money-market and inverse measures. The authors assume deterministic risk-free rates and use the USD money market account as well as the option's underlying price as two numéraires. Moreover, in line with the market completeness assumption as proposed by Jarrow and Larsson (2012), and assuming equivalence of both martingale measures induced by the numéraires, Sepp and Rakhmonov (2022) demonstrate the same relationship between direct and inverse options as Lucic (2022). By the same token, they extend their methodology and consider futures contracts as underlying and illustrate an equivalent relationship between inverse and direct options. While the primary focus of Sepp and Rakhmonov (2022) is not necessarily the valuation of inverse options, but rather the introduction of a novel log-normal stochastic volatility model, the authors nevertheless derive a solution for the joint valuation problem of direct and inverse options applicable to stochastic volatility models. They demonstrate that the put-call parity and duality conditions of FX options remain valid for inverse options under their proposed log-normal stochastic volatility model with quadratic drift. Additionally, they derive the corresponding pricing formulae for direct and inverse options and the necessary and sufficient conditions under which price and inverse price processes are martingales with finite moments, for the log-normal SV models with quadratic and linear drifts, the Heston model, and the exponential Ornstein-Uhlenbeck model. Similar to Lucic (2022), the work of Sepp and Rakhmonov (2022) is primarily theoretical in nature and does not take into account real market conditions, such as market incompleteness or the implications of crypto-denominated profits for traders.

In accordance with the findings of the two papers, the pricing and evaluation process can be done in both USD and bitcoin. Indeed, prevailing academic consensus considers only the USD value of bitcoin options and their profits. An extensive array of finance literature has focused on various stochastic volatility or generalized auto-regressive conditional heteroskedasticity (GARCH) models as mechanisms for pricing bitcoin options (Chi and Hao, 2021; Hung et al., 2020; Siu and Elliott, 2021; Hilliard and Ngo, 2022).[1] For instance, Madan et al. (2019) apply

---

[1]Innovative methodologies incorporating neural network pricing have also been explored (Pagnottoni, 2019;

and test an assortment of models empirically, including the BS model, Laplace model, variance gamma (VG) model, bilateral (double) gamma model, VG Sato model, VG-CIR model and Heston (1993) model, within the context of the bitcoin market. The authors compare the fits based int the Root Mean Squared Error (RMSE), average absolute error (AAE) and the average pricing error (APE). The findings reveal that the BS model captures the surface inadequately, while more sophisticated models yield superior fit. In a significant contribution to the field, Hou et al. (2020) present an essential set of results, highlighting the importance of jumps and co-jumps. They propose a stochastic volatility with a correlated jump model for bitcoin option pricing. Diverse volatility model's accuracy of option prices are investigated by modelling the dynamics of bitcoin with a stochastic volatility model (Heston (1993)), a SV with jumps in the returns (SVJ, Bates (1996)), a SV with contemporaneous jumps in returns and volatility (SVCJ, Duffie et al. (2000)) and a SV with nonlinearity structure (BR, Bandi and Reno (2016); and estimate the diffusion term using the Markov Chain Monte Carlo method. In particular, these models are useful to price exotic options, e.g. cliquet or ratchet options. The authors show that the SVCJ and BR model can describe the log-return dynamics of bitcoin well and use the Crude Monte Carlo technique to approximate prices of options with maturities between one day and two years, and strike levels 50% below and above the assumed underlying price, using the estimated parameters for the different SV models. In a related vein, Cao and Celik (2021) propose a mixed jump-diffusion SV model, assuming a constant interest rate and jump-diffusion for the underlying currency process. On the other side of the spectrum, Cretarola et al. (2020) introduce a bivariate model for the bitcoin price dynamics which is driven by market attention and sentiment. Here, the GBM assumption of the underlying price dynamic is extended by an exogenous factor representing investor's attention in the bitcoin market. The authors show their model is arbitrage-free and introduce a quasi-closed form for pricing European bitcoin options using the Profile Quasi Maximum Likelihood method. In another another strain of literature, (Jalan et al., 2021) compare simulated bitcoin option premia and Greeks with widely recognised valuation models, such as BS and Heston and Nandi (2000) GARCH(1,1) asymmetric multiple volatility regime model. The authors allow for two different volatility regimes and find that the Heston and Nandi (2000) GARCH model models option prices closer to market prices.

All cited scholarly works underscore the vast potential applicability of their respective models, positing the consideration of key attributes intrinsic to bitcoin, such as jumps, and promise success in the area of pricing and hedging. Yet, none of the papers mentioned above, with the exception of Madan et al. (2019), Cretarola et al. (2020) and Jalan et al. (2021), test their models on an empirical dataset, let alone evaluate their performance in a delta-hedging study. Madan et al. (2019) analyse bitcoin options from the LedgerX exchange between July and August 2018. At the time of writing, the LedgerX exchange has ceased operations, following a shutdown in 2021 and subsequent acquisition by FTX. During the short duration

---

Li et al., 2019) but the literature on this strain of research is rather thin.

of 2018, LedgerX was characterised by minimal trading volume, and the bitcoin market has undergone significant transformation since. Therefore, the results derived from that limited time period may deviate substantially when considered in the context of more liquid and current data. Similar, Cretarola et al. (2020) focus on bid- and ask-prices of bitcoin options on Deribit in July 2017. It should be noted, however, that Deribit exchange just started their operations during the authors' sample period, leading to a very low average trading volumes and negligible OTM option trading volume. Given the extensive changes in the bitcoin (and its options) market since then, the influence of these findings may not be as pertinent today, and the methodological approaches require validation with more recent data. In fact, a study by Jalan et al. (2021) demonstrates that the general volatility landscape of bitcoin underwent significant shifts following the introduction of bitcoin futures by the CME in 2019. Consequently, results considering options and volatility prior to the introduction of futures necessitate updating with more current data. Similarly, Jalan et al. (2021) limit their research on the option chains of six available maturities at the evaluation date (27 January 2020) and simulate option prices up to the maturity using a simple Monte Carlo approach. While the authors argue that liquidity and trading volume in January 2020 is sufficient, relying solely on the results of a single snapshot limits the explanatory power of their findings. Furthermore, they fall short in providing a practical implementation of their Greeks discussion and fail to demonstrate their hedging performance.

2.2 Hedging

A great deal of using SV models to price options in the first place is to derive accurate hedge ratios, particularly the price sensitivity, i.e. delta. It is indisputably that the benchmark for any study of dynamic delta-hedging is the Black and Scholes (1973) model. The BS delta only requires a partial derivative of the model option price with respect to the underlying price, because the model assumes a zero correlation between the underlying price and its volatility. But it is well known that equity index options have a large and negative spot-volatility correlation which leads to a pronounced skew in the implied volatility curve, or implied volatility smile, see for example Bouchaud et al. (2001).This empirical observation consequently raises the question: is it possible to beat the BS model in terms of delta-hedging efficiency?

2.2.1 Model-Dependent

Empirical studies focusing on hedging bitcoin options remain limited in the current literature. Teng and Härdle (2022) calibrate multiple stochastic volatility models to Deribit inverse options and perform a dynamic delta-hedging study using an array of nested model-dependent deltas. Specifically, the authors consider deltas of BS, SV, SVJ and SVCJ models, similar to Hou et al. (2020), and measure out-of-sample performance from December 2021 to February

2022, based on the hedging error and benchmark it to the BS delta. Motivated by liquidity and trading clusters, the authors segregate the sample into OTM put, ATM and OTM call options, based on their moneyness. These options have expiry dates up to five days, between five days and four weeks, and longer then four weeks. The authors' approach to dynamic delta-hedge is straight forward, including a self-rebalancing portfolio with initial value of zero. This involves writing an inverse option and taking a long position in the spot market, sized according to the delta. Any excess amount is invested in a risk-free money market. This position is then rebalanced on a daily basis until expiry of the option. Teng and Härdle (2022) find that it is difficult to identify one single model that consistently outperforms the BS delta. Particularly for short- and mid-term options, the authors find almost indistinguishable results, while the longer termed options show mixed results. They also conclude that the broader SVCJ model outperforms its nested models in terms of in-sample and out-of-sample pricing.

Indeed, Matic et al. (2023) offer one of the few detailed studies of bitcoin option hedging. The authors conduct a comparative study of the delta-gamma-vega hedging performance across various stochastic volatility and jump diffusion market models for a multi-asset portfolio. For this, they use daily implied volatilities derived from option mark prices quoted on the Deribit exchange between April 2019 and June 2020.[2] These are then used to calibrate the parametric stochastic volatility inspired (SVI) model, and subsequently interpolate the implied volatilities of options between one and three months in an arbitrage-free way. The study is divided into three distinct sub-periods to account for the prevailing market regime, namely a bullish, a calm and a COVID-phase period. The underlying bitcoin prices are simulated using the SVCJ process introduced by Duffie et al. (2000) with a GARCH-filtered kernel density estimation proposed by McNeil and Frey (2000). The latter technique, being non-parametric, is deemed to closely reflect traded market data. Both models are calibrated using closing prices for bitcoin during the sample period, although the authors do not specify the exact closing time for bitcoin. By doing so, the authors claim to cover a wide range of market dynamics. Subsequently, they compare the performance of the BS Greeks. i.e. delta-, delta-gamma, delta-vega, as well as minimum variance approach, with those derived from a wide variety of stochastic volatility jump diffusion models. This includes comparisons with the hedging performances of a BS, Merton (1976) jump diffusion, SV, SVJ, SVCJ, VG and CGMY-model (Carr et al., 2002). Parameters for each model are computed using a fast Fourier transformation method, introduced by Carr and Madan (1999). The dynamic hedging setup employed Matic et al. (2023) is similar to that of Teng and Härdle (2022) – at the start of each pre-defined market period, the authors write a European option with a maturity of either one or three months and dynamically hedge the exposure of each option. For a delta-gamma or delta-vega hedge, an additional option is introduced to neutralise the

---

[2]There is a slight but important difference between mark and market prices. While market prices are the current bid, ask and mid prices in the options orderbook, the mark price as a 'fair value' estimate of the exchange. How exactly the exchange calculates this value is not stated online, not it is replicable, but it does not necessarily mean that mark and market prices are identical. In fact, for almost any bitcoin option, the market price deviates form the mark value.

gamma or vega exposure. The performance of these hedging strategies is evaluated by measuring the Profit and Loss (PnL) relative to the initial option price, the standard deviation of the hedging error and evaluate the tail behaviours by the expected shortfall of the PnL. Matic et al. (2023) distil their findings into three key conclusions. Firstly, multi-instrument hedging, such as delta-gamma or delta-vega, considerably outperforms simple delta-hedges, particularly with regards to tail-end risk. Secondly, echoing the findings of Teng and Härdle (2022), they find little improvement over simple BS hedges for options with a one-month expiry. Lastly, they note that more complex models significantly enhance hedging performance for options with a three-month maturity.

To the best of our knowledge, these are the only paper considering an active model-dependent dynamic delta-hedging study of bitcoin options. Notwithstanding the comprehensive insights and valuable contributions made by these studies concerning the hedging behaviours of bitcoin options, their methodology incorporates the simulation of market conditions which may skews the real-world use of the results. Moreover, the real-time simulation of parameters requires considerable computational power. Given the recent surge in trading activity and liquidity, it is both feasible and required that future research in this field considers adopting model-free methodologies.

2.2.2 Model-Free

Bates (2005) highlights the disadvantages of computing deltas from parametric models and points out that for a broad class of option pricing models, the appropriate deltas for hedging option positions can be inferred directly from the implied volatility smile. However, the application of such a model-free hedging approach in the realm of cryptocurrency options appears to be unexplored in current literature. This smile-implied model operates under the sole assumption that the underlying is a scale-invariant process.[3] The author indicates that an option delta can be inferred direction from its sensitivity to the strike level using Euler's theorem for homogeneous functions. This process involves calculating the BS delta and subsequently deducting the product of the BS vega and the slope of the volatility smile with respect to the strike level, which can be evaluated numerically (Shimko, 1993). Bates (2005) further points out that the volatility smile, at least for equities, exhibits a downward slope for low strike prices, implying that BS deltas understate the true delta of low-strike options. Alexander and Nogueira (2007a) provide more general results to Bates (2005) ideas. They show that most stochastic processes used for option pricing are scale-invariant and it is possible to use the slope of the implied volatility curve to imply an adjustment to the BS delta which is model-free, in the sense that it is the same for any scale-invariant model. They conclude that any difference between the empirical hedging performance of two parametric volatility models for tradable instrument only arises because of different calibration errors.

---

[3]That is, the price shows constant returns to scale, making option prices a homogeneous functions of degree one with respect to the underlying price and the strike level.

Another approach in adjusting the BS delta by adding a term which captures the spot-volatility correlation is to use the smile-adjusted deltas proposed in the pioneering work of Derman and Kani (1994), Rubinstein (1994), Dupire (1994) and Derman et al. (1996). These are not exactly model-free, because the adjustment term depends on a parameterisation of local volatility which itself depends on the prevailing regime of the market. However, they are model-free in the sense that there is no process, such as a stochastic local volatility jump diffusion, which is assumed to drive the underlying price evolution, and no parameters to calibrate using option price and/or underlying historical data. A great practical advantage of these methodologies is that deltas are easy to compute and are implemented in a cost-efficient way. There is no requirement for model calibrations because all information is derived from the market implied volatility smile in a straightforward and robust, model-free manner.

Derman and Kani (1994) and Rubinstein (1994) show how to use the implied volatility smile as input to deduce the underlying's dynamics. More specifically, the authors show how to extract a unique binomial tree, in line with Cox et al. (1979) but non-constant volatilities at each node, for the underlying corresponding to its true future evolution. The general approach first estimates the risk-neutral probability distribution of underlying at the end of the tree, and then determines the up and down step sizes and probabilities throughout the tree that are consistent with the implied probability distribution. In this context, both Derman and Kani (1994) and Rubinstein (1994) are among the first to introduce the concept of local volatility. In contrast to Black and Scholes (1973), the local volatility is the instantaneous volatility and a function of both time and underlying price that "locks in" the forward volatilities. Alexander (2008) uses this notation and describes the forward volatility as the volatility at some future point if the underlying takes the a certain values and draws an analogy to the forward rates in bond markets. In essence, it reflects the short-term ATM implied volatility at the underlying level. In fact, Derman et al. (1996) provide an extensive comparison between the local volatility and forward rates and motivate the use of local volatility to price exotic options. Derman and Kani (1994) provide a discrete algorithm to calculate index levels, transition probabilities and Arrow-Debreu prices at each node iterative level by level.[4] At the same time, Dupire (1994) was addressing the same issue, leading to a equation that allows the local volatility to be calculated continuously from current option market prices. This proposed equation relies solely on option prices, strike levels, and partial derivatives of the option chains with respect to maturity and strike, thereby making it immune to changes in the underlying asset. That is, the local volatility of Dupire (1994) remains unaffected for price movements while the BS IVS changes. This characteristic of the local volatility is commonly know as the static smile property, while the other refers to the floating smile property.

Derman et al. (1996) extend the work of Derman and Kani (1994) and emphasise the use of local volatility surfaces to price exotic options. Once fitted, the implied tree model produces a grid of future underling prices, each corresponding to a specific local volatility. Assuming a

---

[4]The Arrow-Debreu price is the price of an indicator option that pays 1 unit payoff in only one state.

constant market perception of local volatility over time, they demonstrate how local volatilities can be deployed to ascertain the relationship between implied volatility and strike price in the future. This methodology uses the tree's directional characteristics to compute fair values for options across all strikes and maturities at a future time. The conversion of these prices into BS implied volatilities enables the computation of equitable future implied volatility surfaces. Derman et al. (1996) propose three heuristic rules, which were subsequently generalised and further examined by Kani et al. (1996). Particularly the third rule is of great interest as it considers the correlation between underlying and implied volatility. The authors provide an intuitive explanation of the implied volatility as a global average over local volatilities. Assuming that that local volatility is a linear function of the underlying price, independent of future time, the authors derive a smile-adjusted delta ratio accounting for the BS delta, vega, and observed sensitivity of implied volatility to strike level. Specifically, for equity indices, the implied volatility is inversely correlated with the index level. They demonstrate that, in the implied tree model with negative skew, the exposure of both call and put options is consequently lower than it would have been in a BS environment with flat volatility.

Derman (1999) focuses on observed sensitivities of implied volatility to underlying and identifies three main aspects of the skew that are invariant, or 'sticky', as the underlying moves. His main motivation emerges from empirical observation or the S&P 500 index and the corresponding three month implied volatilities. He identifies distinct volatility behaviour for certain market periods, or regimes, and introduces the famous sticky models, i.e. sticky strike, sticky moneyness (or delta) and sticky tree. Each of these models effectively represents the volatility-spot relationship for its respective market regime:

Sticky Strike

The Sticky Strike (SS) model disregards any correlation between the underlying and implied volatility and depicts a stable-trending market. Given the current skew, some traders believe that, as the index progresses in any given direction, the volatility of an option with a specific strike and maturity remains unaltered; hence its name. If one were to visualise the implied tree, each option would have its own tree, each with a distinct constant volatility equal to its own implied volatility. Derman (1999) himself refers to this approach as "a poor man's attempt to preserve the Black and Scholes (1973) model" as it provides each option an independent tree, without concern for the collective consistency of the options market's view of the index. Under the sticky strike assumption, the Black and Scholes (1973) delta is not adjusted to account for changes in implied volatility as spot price fluctuations occur.

Sticky Moneyness/Delta

In the Sticky Money (SM) or sticky delta framework, the implied volatility of an option depends only on its moneyness, or its delta, and describes a range-bounded market.[5] Any variation on the volatility stems entirely form its dependence on the single variable moneyness. The intuition behind this model is that ATM implied volatility should remain the same, independent on the current underlying level or its changes. That is, as the implied volatilities for – theoretically – all moneyness/delta are fixed, the ATM implied volatility shifts to the ITM/OTM implied volatility as the underlying moves. This holds for ITM/OTM options as well, i.e. the option that is 20% out of the money after the underlying moves should have the same implied volatility as the 20% OTM option before the index move. The implication of this assumption for the delta is significant. Within the sticky moneyness assumption, if the spot price of the underlying asset changes, the delta of the option would change accordingly to keep the level of moneyness constant. Derman (1999) shows that the SM delta adjustmentment is greater than the BS delta.

Sticky Tree

The Sticky Tree (ST) model completes Derman's sticky models and describes a falling crash-jump market with a lot of fear and uncertainty. This approach considers the implied tree (Derman and Kani, 1994; Derman et al., 1996) of instantaneous volatility of future underlying prices derived from current market implied volatility surface. Instead of allocating a separate tree to each individual option or moneyness level, all options are subject to the same implied tree and move across its nodes in tandem with alterations in the underlying. This suggests a distinctive pattern where implied volatilities decline in response to increases in either the strike or the underlying. Here, ATM implied volatilities decrease twice as much with the underlying. Furthermore, given that volatility decreases as the underlying increases (at least in the case of equities), the option's delta is found to be smaller than its BS counterpart. In a sense, it is the only true local volatility model because it keeps the floating properties of Dupire (1994) and considers only one implied tree as proposed by Derman et al. (1996).

Apart from the smile-implied and -adjusted deltas, there are more generic deltas which minimise the instantaneous variance, i.e. the Minimum Variance (MV) delta. Bakshi et al. (1997) are among the first to introduce another model-free delta adjustment which minimises the variance of the hedging error of a delta-hedged options position. As shown by Alexander and Nogueira (2007a), the minimum variance total derivative with respect to price is another delta which accounts for non-zero spot-volatility correlation, but it is model dependent. Nevertheless these authors cannot distinguish the empirical results obtained using the model-free MV delta of Lee (2001) and those MV deltas based on different scale-invariant models. This delta is also 'smile-adjusted', in the sense that it adds a term to the BS delta that is calibrated

---

[5]In the BS model, the delta itself depends on the underlying and strike through the moneyness variable, so that "sticky moneyness" is equivalent to "sticky delta,".

using the empirical characteristics of the implied volatility smile curve. The author provides an extensive review of stochastic (local) volatility option pricing models and generalises and extends the minimum variance hedge ratio of Bakshi et al. (1997) in which her proposes an BS delta adjustment similar to the sticky moneyness model of Derman (1999).

Hull and White (2017) put forth a distinctive methodology for estimating the MV delta using a different estimation of the vol-spot sensitivity. Similar to Alexander et al. (2012), they test the effectiveness of an empirical estimate of the sensitivity for delta-hedged options on the S&P 500, S&P 100, the Dow Jones Industrial Average between January 2004 and August 2015. For this, they consider the BS delta and estimate (forward-looking) the minimum-variance delta which yields a hedging error of zero. The researchers find that the BS delta is consistently higher than the MV delta and has a roughly quadratic relationship to the BS delta, independent of the options maturity or strike. Drawing from these findings, the authors consider the smile-adjusted delta, obtained via the application of the chain rule in the manner of Lee (2001), and characterise the spot-volatility sensitivity as a quadratic polynomial function of the BS delta. To carry out this analysis, the authors adopt a rolling window spanning three years to conduct a regression of the instantaneous hedging error on the squared BS delta. The outcomes of their study are somewhat mixed. Hedging operations prove to be more effective for call options than put options, and more favourable for OTM options compared to ATM options. Nonetheless, a dynamic delta-hedging comparison with alternate models demonstrates their model excels over its (SV) contemporaries.

Despite all the presented advantages of SV models, Alexander and Nogueira (2007b) show that *every* stochastic and/or local volatility equity option pricing model, for a *tradable* instrument but not an interest rate, should fall into the scale-invariant class, however complex the additional features such as jumps or Lévy processes. Consequently, any difference between the empirical hedging performance of two parametric volatility models (for a tradable instrument) only arises because the models have different calibration errors. The delta (and indeed the gamma) partial derivatives of the option price with respect to a tradable instrument price are theoretically identical to the model-free scale-invariant delta. By contrast, the delta derived from a non-scale-invariant model, such as the local volatility model of Dupire (1994), or the sticky-tree model of Derman and Kani (1994), is not theoretically identical to the scale-invariant delta. Neither is a minimum-variance delta, which is the total derivative that includes the vega effect arising from a non-zero spot-volatility correlation. Moreover, the simple scale-invariant delta derived by Bates (2005) is greater than (less than) the BS delta when the slope of the smile is negative (positive). Since Coleman et al. (2001) show that the BS delta tends to over-hedge in a local volatility framework, scale-invariant deltas will over-hedge even more than the BS delta when the implied volatility skew is negative

There are several previous empirical studies of smile-implied and/or smile-adjusted delta-hedging, but all of them study equity index options. Not all of the results are consistent: Vähämaa (2004) shows that some smile-adjusted deltas out-perform the BS delta for FTSE

100 index options, but only during excessively volatile periods; Crépey (2004) confirms these findings for DAX 30 options; Attie (2017) claims that smile-implied deltas consistently out-perform the BS delta for hedging S&P 500 options; Alexander et al. (2012) extend the Derman (1999) framework to a Markov-switching setting which reflects the correct smile-adjusted delta for the prevalent market regime, showing that, for S&P 500 options, it is only possible to improve on the BS delta by using this Markov-switching extension; and François and Stentoft (2021) also examine S&P 500 options and confirm that standard adjustments cannot out-perform the BS delta or delta-gamma hedges, but their new smile-implied delta-gamma-vega hedge substantially improves on the BS model. Much less is known about the success of smile-adjusted delta-hedges for other types of options. In this strand of the literature, Nastasi et al. (2020) calibrate smile-consistent models for commodity options to capture the smile dynamics and Malz (2000) explains how to take smile adjustments into account when measuring the risk of foreign exchange options. None of these models have been applied and tested on cryptocurrency option markets.

2.3 Volatility Trading

The implementation of a delta-hedging strategy on an options position primarily serves to mitigate underlying risk. It is, however, the most simple form of pure volatility trading as the delta-hedge removes all the exposure to the underlying. While volatility as an own asset class on other assets like equity or commodities has been present and popular for decades (Derman et al., 1996), the notion of engaging in pure cryptocurrency volatility trading remains in the early stages of its development. Nevertheless, interest in this area is increasing, as affirmed by recent studies from Aalborg et al. (2019) and Chen and Huang (2021). Carr and Lee (2009b) provide an excellent summary of the history of pure volatility trading, starting with the first variance derivatives issued by UBS in 1993, to the modern volatility derivatives traders use today.[6]. Carr and Madan (2001) provide practical insights of the volatility trading history and present specific positions where it is feasible to engage in volatility trading, for instance, through static delta-hedged options, variance swaps, or futures on volatility indices.

---

[6]The inception of volatility trading can be linked back to the seminal works of Neuberger (1994) and Dupire (1993). Neuberger (1994) presents the concept of the 'log-contract', a financial instrument that yields the log return of the underlying asset. The author underscores how this furnishes an alternative avenue to trade realised volatility, demonstrating within the context of the Black and Scholes (1973) framework, that a portfolio consisting of delta-hedged log contracts on the underlying asset can be effectively utilised to craft a pure volatility strategy. Despite its theoretical appeal, these contracts see infrequent trading due to the complexity inherent in their pricing mechanisms. In a parallel development, Dupire (1993) shows how this payoff structure can be replicated by a static portfolio of options. Additionally, he introduces the notion that a calendar spread of two log-contracts with disparate maturities would yield the realised variance between the two specified dates.

2.3.1   Variance Swaps & Volatility Indices

Variance swaps on equity indices were introduced by Goldman Sachs in the 1990's. Demeterfi
et al. (1999) present the replication strategy, which captures the realised variance over the
lifespan of the swap and emphasise the hedging cost as fair value swap rate. The authors
put forth an intuitive approach wherein they construct a portfolio of options which payoff ap-
proximates the log contract presented by Neuberger (1994), and show re-hedging the portfolio
returns the payoff of a variance swap. Beyond the discrete case where future variances are
averaged, the authors illustrate how continuous rebalancing of a position in the underlying –
such that it maintains a long position on the underlying worth of \$1, combined with a static
short position in a contract that pays the log of the total return – retrieves the cost of replica-
tion directly. Recognising that actively traded log contracts are absent, the authors replicate
the log payoff at all underlying levels at expiration by decomposing its shape into linear and
concave components, considering each component separately, as proposed by Dupire (1993).
The first component, linear, is replaced by a forward contract on the underlying asset, with
the same expiration as the swap. The second component, curved – accounting for quadratic
and higher order contributions – can be replicated using standard options across all possible
strike levels with the same maturity. Finally, the authors present the decomposition of a
log payoff as a portfolio comprised of a short position in forward contracts, a long position
in put options – each inversely weighted by their squared strike value, for all strikes up to
the underlying value; and a long position in call options, inversely weighted by their squared
strike, for all available OTM strikes. The corresponding fair-value variance swap rate formula
is approximated as a Riemann sum over these discrete and bounded option positions.

The authors draw attention to a series of intrinsic limitations of their replication method:
(i) the use of discrete strike levels for options causes sparse replication, leading to the dis-
cretisation error; (ii) the unrealistic assumption of the strike ranges (truncation error) leads
to underestimation of the tails; and (iii), the occurrence of jumps and resulting jump error
neglects its impact on the volatility. In fact, both truncation error (Jiang and Tian, 2005) and
discretisation error (Bernard and Cui, 2014) affect its accuracy significantly. The jump error
can be quite large during volatile periods, see Aït-Sahalia (2004) or Rompolis and Tzavalis
(2017). For instance, during the financial crisis of 2008 the VIX commonly underestimated
the S&P 500 variance swap rate traded in over-the-counter (OTC) markets by 5% or more.

It is worth noting that Demeterfi et al. (1999) assume the underlying follows a geometric
Brownian motion. Bakshi et al. (2003) however, contest this assumption, showing that there
is no necessity for the underlying asset to adhere to a specific dynamic, provided it is a
positive semi-martingale. The authors further present an alternate fair swap rate which takes
into consideration the second moment of log returns as opposed to the expected value of the
quadratic log returns. In this vein, both Jiang and Tian (2005) and Carr and Lee (2009a)
generalise the geometric diffusion assumption of Demeterfi et al. (1999) and demonstrate that
the variance swap replication still holds under any diffusion for the stochastic volatility process,

including non-zero spot-volatility correlations. Furthermore, Gatheral (2006) introduces an alternative methodology, validating the swap replication of Demeterfi et al. (1999) via a model-free decomposition of a power option. In this approach, he computes the first moment of the quadratic variation of the returns process, assuming a zero spot-volatility correlation.

Modern implied volatility indices are build upon the work of Demeterfi et al. (1999). The first idea of a volatility index goes back to Brenner and Galai (1989) who propose financial instruments for hedging volatility, similar in spirit to Neuberger (1994). The CBOE, in consultation with Whaley (2000), introduced a index of implied volatility on the S&P 100 index and only using at-the-money options. The CBOE subsequently updated their volatility index methodology (CBOE Global Markets, 2003), drawing upon the academic works of Breeden and Litzenberger (1978); Madan et al. (1998); Demeterfi et al. (1999); Britten-Jones and Neuberger (2000). This marked the advent of the VIX derived as a corresponding fair-value swap rate for the S&P 500, informally referred to as 'The Investor Fear Gauge'. This variance-swap-based volatility index allows for the determination of the (ex-ante) fair swap level. Given it represents the expected realised variance, it conveys that the swap rate is the market-implied future variance. In 2004, the CBOE began listing VIX futures, with the corresponding European options introduced in 2006. The trading of futures and options on the VIX has become widely established, albeit highly speculative. Subsequently, implied volatility indices on other equity indices like the DAX 40, or alternative asset classes such as commodities or FX rates, have increased rapidly, providing a foundation for futures trading.

As already mentioned briefly, the methodology employed by the CBOE is criticised for it vulnerability to manipulation, as well as its dismissal of jump and discretisation errors. This perspective is supported by both theoretical and practical literature (Bennett, 2014; Griffin and Shams, 2018) which demonstrates how such calculations may be susceptible to manipulation during the settlement of indices and futures. In the context of bitcoin, this errors are intensified, as presented by numerous studies analysing jumps in bitcoin prices. Scaillet et al. (2020) employ the detection method of Lee and Mykland (2008) on daily data to detect jumps between June 2011 and November 2013 using the database leak of the Mt.Gox exchange. They find approximately one jump per week in the BTCUSD rate and claim that order flow imbalance and the aggressive traders, as well as a widening of the bid-ask spread predict the jumps. However, their sample covered a period when the bitcoin universe could almost be described as prehistoric with two major crashes caused by fraud and manipulation (Cheung et al., 2015; Gandal et al., 2018). The market has matured considerably since then, with initial coin offerings of almost 4,000 new crypto assets starting in 2017 and continually growing, along with the introduction of bitcoin derivatives. Research by Hilliard and Ngo (2022) find similar results for more recent data and analyse the impact of these jumps on option pricing. Kang and Kim (2019) and Hu et al. (2019) use the methodologies introduced by Jiang and Oomen (2008), Barndorff-Nielsen and Shephard (2004) and Barndorff-Nielsen and Shephard (2006) to detect and analyse more recent jumps in bitcoin prices, between 2014

and 2018. Despite identifying more jumps in bitcoin as compared to the S&P 500, they found that these jumps accounted for a smaller proportion of the total return than its continuous component, an outcome attributable to the high volatility of bitcoin prices.[7]

In light of the limitations identified in the variance swap based calculation, alternative model methodologies have been proposed. Chow et al. (2021) demonstrate that the VIX tends to underestimate volatility in scenarios of negative return skewness, attributing this to the discretisation error inherent in the CBOE methodology. They introduce a Generalised VIX (GVIX) methodology which eliminates the necessity for a diffusion assumption. However, the authors do not provide empirical evidence validating their claim of superiority of their methodology over the VIX in terms of forecasting realised variance. Leontsinis and Alexander (2017) put forward a formula for a fair-value of realised variance, where it is defined as the average sum of squared price changes, as opposed to the change in log prices. Their rationale is grounded in the employment of a definition that adheres to the aggregation property, as introduced by Neuberger (2012) and Bondarenko (2014). Indeed, their arithmetic variance swap bears a striking resemblance to the simple variance swap posited by Martin (2013), wherein realised variance is defined as the average sum of squared returns. Nevertheless, the simple variance swap is only applicable to underlyings with positive prices and fails to satisfy the aggregation property, leading to a non-zero discrete monitoring error. A distinctive advantage of the realised variance definition proposed by Leontsinis and Alexander (2017) is that the fair-values exhibit no jump bias. Furthermore, their arithmetic index assigns a weight to each option based on the squared underlying values, rather than the squared strike. By contrast, the standard 'geometric' formula adopted by the CBOE can produce substantial errors in the presence of jumps. Unlike Chow et al. (2021), Leontsinis and Alexander (2017) conducted empirical tests of their model, finding that their arithmetic approach surpassed the geometric VIX index in predicting realised variance during the 2008 banking crisis.

The development of volatility indices for cryptocurrencies remains largely under-explored. Among the few noteworthy alternative volatility indices is the VCRIX, an implied volatility index pertaining to the CRIX index, initially introduced by Kim et al. (2021). The authors claim that the CRIX serves as a unique index for the crypto-market, mirroring the role of the S&P 500 in the traditional financial landscape. Owing to the absence of a derivatives market for the CRIX, the VCRIX does not adopt the swap-based options methodology propounded by the CBOE Global Markets (2003). Instead, the authors aim to uncover a model that embodies

---

[7]Another strain of research (Qian et al., 2022) focus on the role of jumps in forecasting bitcoin realised volatility using linear and non-linear models. Their results offer substantial evidence that a forecasting model incorporating continuous-time jumps and two-stage regimes can markedly enhance predictive accuracy and realise significant economic gains. They also note that the superior predictive capabilities of the model are most pronounced during periods of high volatility, particularly during Black Swan events. Gkillas et al. (2022) look into the analysis of jump size and direction, proposing an asymmetric jump model. Their research indicates that downside, upside, and small jumps in cryptocurrencies have a negative effect on the jump component of other cryptocurrencies' realised volatility, whereas large jumps produce the contrary effect. They present evidence of co-jumps, substantiating the findings of Bouri et al. (2020). Overall, the literature indicates that while jumps are present and do not significantly contribute to the total return, they tend to occur in clusters, which may influence risk premia over longer maturities.

the predictive potency equal with traditional implied volatility indices. Consequently, the VCIRX serves only as a proxy for volatility, chosen from a selection of various volatility models in a comparative examination. The authors use the annualised historical rolling volatility of log returns over a 30-day period and the realised volatility derived from intra-day observations of the underlying as two surrogates for implied volatility. These proxies are forecasted using an assortment of GARCH models, the Heterogeneous Auto-Regressive (HAR) model of Corsi (2009), a Long-Short Term Memory (LSTM) model and various multivariate GARCH models. The analysis is conducted on daily closing price data of the CRIX from November 2014 to January 2021, with each model being evaluated over the period June 2016 to May 2020 based on the mean absolute error and mean squared error. Notwithstanding, a comparison to a naive model and assessment of the directional error would have potentially contributed valuable insights. The authors conclude that the HAR model outperforms the other contestants. To validate their choice of proxies, Kim et al. (2021) construct a version of the VIX (named the approximated VIX, or AVIX) which uses the selected HAR model to process the log-returns of the S&P 500, as opposed to the CRIX. From the S&P 500 log-returns, they derive a 30-day rolling historical volatility and formulate the AVIX. They subsequently compared the AVIX to the actual values of the CBOE VIX, arguing that this would establish the 30-day rolling historical volatility and HAR model as a suitable surrogate for a volatility index. The authors assert that they had successfully developed an effective forward-looking implied volatility index for the cryptocurrency market. However, it is important to note that neither the CRIX nor its VCRIX, nor any proposed ETFs on the indices, have seen commercial implementation.

Venter and Mare (2020) propose a GARCH volatility index wherein the expected volatility value represents the the anticipated arithmetic mean of the variance for the forthcoming 30 calendar days. To achieve this, the authors model the dynamics of the underlying bitcoin price using a square-root stochastic autoregressive volatility model (SR-SARV, Meddahi and Renault (2004)). Using daily bitcoin and CRIX data spanning from January 2016 to 2019, the authors calibrate the dynamics and report the implied volatility index for varying maturities. Echoing the approach of Kim et al. (2021), this index is derived from modelling the underlying rather than extracting the expected volatility from the options market. In contrast, Woebbeking (2021) presents the only genuine implied volatility index, in that its computation is grounded on a fair variance swap rate. The author compares the implied volatility index methodology of Demeterfi et al. (1999)with the interpolated 30-day ATM implied volatility deduced using the Black (1976) model. To facilitate this comparison, the author considers options and futures data from Deribit spanning February 2020 to July 2021 and contrasts both the bitcoin volatility index and IV with those of other asset classes. The analysis reveals that the index and implied volatility frequently manifest an inverse relationship to their underlying, a finding that is typical in such studies. Woebbeking (2021) concludes that the swap-based computation provides a more accurate portrayal of market expected volatility.

2.3.2 Variance Risk Premium

The variance risk premium may be defined as the reward required by a risk-averse investor to compensate for the risk of both stochastic volatility and jumps in price of the underlying asset, see Todorov (2010) and Konstantinidi and Skiadopoulos (2016a) for example. Two distinct methodologies are used to assess the VRP: The 'model-free' approach, which identifies the VRP with the expected PnL on a swap that exchanges realised variance for a fixed swap rate; and a model-dependent approach, in which both underlying and option prices are modeled using common stochastic volatility models. The first strand of literature for the former approach takes its roots from the identification by Bakshi and Kapadia (2003) of the VRP with delta-hedged gains on options and the subsequent work of Bakshi and Madan (2006), which links the VRP with higher moments of returns. The latter strand of the theoretical literature, following Broadie et al. (2007), calibrates a stochastic volatility-jump model using both option prices and the underlying time series. However, this approach depends on the specification of the model for the underlying price and volatility processes (Egloff et al., 2010).

The literature on cryptocurrency variance swaps and resulting risk premiums is very limited. However, a substantial body of empirical research literature analyzes the variance risk premiums in different asset classes. Among these, equity variance swaps represent the majority of variance swaps being traded. For instance, in their study, Carr and Wu (2009) find that the S&P 500 risk premium is usually negative because investors are willing to pay for the strong diversification offered by an instrument that maintains such high negative correlation with US equity returns. The authors estimate the value of the 30-day variance swap rate on individual stocks and stock indices by considering a portfolio of vanilla options, subsequently comparing the synthetic 30-day variance swap rates to the ex-post realised variance to ascertain the variance risk premium, which corresponds to the payoff of a long position in a variance swap. This investigation assumes ideal market conditions, that is, a frictionless market without jumps where continuous monitoring of the variance is feasible. It operates within the model-free environment as proposed by Carr and Madan (2001). The empirical evaluation of this study addresses approximation errors and scrutinises the impact of the option's bid-ask spread on the VRP. The sample, which consists of option closing prices of five stock indices and 35 individual stocks spanning from January 1996 to February 2003, reveals a marginally negative VRP for the S&P 500 and Dow Jones Industrial Average index with sudden and strong upward jumps in periods of high volatility.

Todorov (2010) pioneers the effects of jumps on the variance risk premium. Implementing a broader jump framework, he finds that investor risk preferences shift towards risk aversion within a jump environment, leading to an elevated VRP. Rather than using the strikes of traded options as a base for his analysis, he employs synthetic strikes calibrated from a semi-parametric stochastic volatility model. Following a similar framework Bollerslev and Todorov (2011) demonstrate that the inclusion of jumps and other uncertainties, such as approximation errors, play a significant role in the pricing of variance swaps. They also investigate the contribution of jump tails, discovering that they directly influence over half the observed VRP, as calculated using S&P 500 option from 1996 to 2008. In fact, Aït-Sahalia et al. (2020) show that the accumulation of these approximation errors can lead to the mispricing of the fair-value swap rates by up to 5% when compared to the market strike. Broadie and Jain (2008b) price variance swaps under a continuous diffusion process, introducing dynamic trading strategies that enable the replication of other volatility derivatives' payoff using a set of European options and variance swap. Their analysis of the replication error reveals a bias resulting from daily variance monitoring as opposed to continuous monitoring. Bernard and Cui (2014) provide an even broader extension, offering analytical formulas for fair variance strike pricing of the underlying with a series of stochastic volatility models.

Aït-Sahalia et al. (2020) provide a model-free term-structure analysis of the S&P 500 VRP. Similar to previous work, they find a general low negative VRP with sudden jumps. The study determines that jumps significantly impact the fair pricing of options, with evidence suggesting that investors demonstrate a propensity to hedge against volatility exposure, particularly in the aftermath of substantial market declines. The authors define the VRP as the objective risk-neutral expectation of future realised variance, using daily traded rates from the period January 1996 to December 2010. The rates are compared with the VIX and the methodology adopted by Carr and Wu (2009), indicating that jumps constitute a significant and time-varying component of the fair-value swap rate. These findings are consistent with those presented by Bollerslev and Todorov (2011) among others. A similar investigation into the predictability of the VRP is conducted by Konstantinidi and Skiadopoulos (2016b). They use S&P 500 variance swap rate data from 1996 to 2009, applying four distinct models to ascertain the predictability of strike rates. Their research reveals that a 'trading-activity model' accurately forecasts the VRP based on both in-sample and out-of-sample backtesting.

Others, such as Bollerslev et al. (2009), seek to predict market returns using the standard 30-day VRP of the S&P 500, with data spanning from 1990 to 2007. They calculate the VRP using the model-free approach, as introduced by Demeterfi et al. (1999) and Britten-Jones and Neuberger (2000), and link the premium to time-varying uncertainties in the market. Their in-sample testing reveals that this provides a robust prediction of aggregate market returns on a quarterly basis. Further exploring the determinants of the VRP, both Bekaert et al. (2013) and Bekaert and Hoerova (2014) assess the accuracy of 14 statistical models designed to forecast realised volatility. They deconstruct the VIX into a conditional variance of the

market and an equity VRP, subsequently using these as input parameters for the models and analysing the generated predictions. The studies yield superior results in terms of accuracy.

Another strand of literature considers the discrete monitoring of variance swaps. Broadie and Jain (2008a) assess the bias induced by jumps under a range of stochastic volatility models and deduce fair-value swap rates for variance swaps monitored discretely. Their strike simulations, based on a selection of parametric pricing models, reveal that discrete sampling and jump error exert a minimal impact on the fair price of variance swaps. Carr and Lee (2009b) demonstrate that the bias stemming from discrete monitoring, as proposed Demeterfi et al. (1999), is typically negative and, similar to the jump bias, is most conspicuous during volatile periods. Jarrow et al. (2013) derive bounds for the error incurred from discrete monitoring and establish that these bounds narrow as the monitoring frequency increases. Bernard et al. (2014) build upon these findings, providing conditions to reduce the bias from discrete monitoring. They scrutinise the assumptions posited by Jarrow et al. (2013) and relax them, subsequently furnishing simpler expressions for the formulas of fair swap rates. Uniquely, Bondarenko (2014) devises a methodology for pricing variance derivatives in an incomplete market and introduces 'generalized' variance derivatives. Such contracts typically employ futures as the underlying asset and compute the realised variance as a sum of squared differences in daily log prices. Contrary to previous works, Bondarenko (2014) assumes that replication is unachievable at all times and accounts for jumps. By incorporating these more realistic adjustments into the pricing formula, he determines that his methodology diverges significantly from their continuously sampled counterparts.

So far we, considered only equity VRP. However, similar results are found in currencies (Ammann and Buesser, 2013), and commodities (Prokopczuk et al., 2017a). Still, most research is on the risk premium for US equities; for example, swaps on commodities and currencies as well as on bonds are traded frequently on OTC markets. Ammann and Buesser (2013) and Domowitz and Hakkio (1985) analyse the VRP for currencies (USD, Euro, Japanese Yen and AUD) finding results similar to those for the equity VRP. Prokopczuk et al. (2017b) give a very detailed VRP analysis of the most traded commodities. Here again, the paper finds similar results compared to the equity VRP as it is slightly negative and jumps in volatile periods. However, some commodities have a low traded volume and the results should be considered with caution. Others like Trolle and Schwartz (2010) analyse the VRP for energy markets (e.g. electricity, gas and heating oil) and Choi et al. (2017) provide a comprehensive VRP analysis for the bond market. With minor deviations, all authors find that the VRP is generally negative.

2.4 Implied Volatility

Implied volatility is often interpreted as traders' forecast of the underlying return volatility over the lifetime of an option. Despite its relevance, the existing literature addressing bitcoin implied volatility remains limited, presumably due to the relatively recent availability of liquid options data – roughly three years.[8] Alexander et al. (2022) examine the behaviour of bitcoin implied volatility smiles of Deribit options to infer whether demand pressures on market makers are motivated by directional or volatility traders. They conclude that volatility traders significantly influence both ATM and OTM option prices, whereas directional trades primarily impact OTM options. Additionally, Zulfiqar and Gulzar (2021) offer a comprehensive overview of the bitcoin options market. Their analysis of the bitcoin implied volatility smile reveals similarites to the skews typically observed in equity markets, especially in the context of the first quarter of 2020.

2.4.1 Modelling

A plethora of different (non-)parametric models have been developed with the objective of capturing the dynamics of implied volatility. The most widely adopted parametric and deterministic models are the BS and local volatility models previously discussed. Dumas et al. (1998) evaluate the implied tree models of Derman and Kani (1994) and Rubinstein (1994) for their effectiveness, introducing an assortment of parametric and semi-parametric deterministic volatility functions. The authors propose volatility as a function of both strike price and Time to Maturity (TTM), putting forth four distinct second-order polynomial functions. Their first model is a simple BS model wherein the volatility remains constant throughout the option's lifespan. The remaining proposals allow for volatility to be linearly and quadratically dependent on both the strike level and time to maturity. To test the fit of their models, they employ conventional statistical measures and assess their delta-hedging performance for S&P 500 options over diverse horizons. The parameters are calibrated weekly, and the performance

---

[8]A robust body of literature has been dedicated to examining the dynamics of realised volatility in cryptocurrencies (Baur and Dimpfl, 2018; Katsiampa, 2019; Ben Cheikh et al., 2020). Although these studies have not directly addressed implied volatility, they all concur on the existence of a positive relationship between return and volatility in the domain of cryptocurrencies. This stands in contrast to equity or commodity markets, where an increase of the underlying typically precipitates a (mild) decrease in volatility. These researchers observe a general trend of increased volatility subsequent to positive shocks. The so-called 'leverage effect', commonly noted in other asset classes, appears to be absent in cryptocurrency markets, as Huang et al. (2022) find. Expanding upon this line of inquiry, Akyildirim et al. (2020) explore the correlation between the realised volatilities of various cryptocurrencies and their relationship with equity volatility indices. Their research provides evidence of substantial correlations with volatility products. In another notable study, (Chaim and Laurini, 2018) delve into the daily returns and volatility of bitcoin, observing that surges in bitcoin returns occur contemporaneously and are associated with significant events, such as hacking incidents and unsuccessful fork attempts. While the literature on realised volatility forecasting is extensive (Shen et al., 2020; Köchling et al., 2020), only Hoang and Baur (2020) stand out as the only researchers to actively consider bitcoin implied volatilities. Using the implied volatility of traded options from Deribit, they seek to enhance forecasts of day/week ahead realised volatility. Their findings indicate that implied volatility presents a more accurate forecast for long-term (10+ days) volatility, while ARCH models exhibit superior performance in the short-term.

of local and BS models is compared for the period from June 1988 to December 1993. The authors conclude that their parsimonious model fits the surface well and that the prediction and hedging error of the fitted model are lower than those yielded from the implied tree model for shorter maturities. Lewis (2000) confirms these findings through a similar study using the second-order polynomial expression of the implied volatility smile to fit and hedge the skew and Coleman et al. (2001) propose an alternative deterministic local volatility model that uses splines to depict the volatility function. This function is derived by resolving a non-linear optimisation problem to fit option prices closely and assess its delta-hedging performance based on 1993 European S&P 500 index options. The authors find that over extended (17+ days) periods, local volatility deltas outperform BS. Both Dumas et al. (1998) and Coleman et al. (2001) are true local volatility functions in a sense that they have the static smile property. Cox (1975) introduces the Constant elasticity of variance (CEV) model which models the price dynamic using only one Brownian motion to drive both price and volatility, i.e. price and volatility are perfectly correlated. Other deterministic models which do not directly model the implied volatility surface but take the information, like the Vanna-Vega model for FX options, require less computational power but are still popular pricing/modelling models among traders. At the time of writing, none of the mentioned deterministic parametric functions have been fitted and tested to cryptocurrency implied volatility surfaces.

Another strand of literature examine parametric models that, nevertheless, exhibit stochastic characteristics.[9] Hagan et al. (2002) propose the Stochastic$-\alpha\beta\rho$ (SABR) model, a natural extension of the CEV model. The SABR model uses the functional form of the CEV to model the dynamics of the forward price with maturity equal to the option's maturity. However, it introduces a stochastic α term, driven by a spot-volatility correlated diffusion process. Gatheral (2006) offers an arbitrage-free calibration approach for the stochastic volatility inspired parameterisation of the volatility smile, which is later broadened to encompass the entire volatility surface (Gatheral and Jacquier, 2014). This five-factor model, premised on the moment formula of Lee (2004), delivers the market implied expectation of volatility. Despite its label, this model is not truly stochastic, in that there are no driving stochastic processes. Instead, it aims to describe surface patterns that would emerge under the assumption of stochastic volatility. The model can be enhanced to an SVI-Jump Wing (SVI-JW) model by incorporating an additional parameter to account for swings in the tails. Schneider and Trojani (2018) present a different approach and show that the surface information is equivalent to a smoother option portfolio surface and emphasise extracting driving factors of the surface directly from the equivalent option portfolio. The authors claim that this has the advantage of providing interpretable factors with implications for trading of nonlinear risks related to the underlying.

---

[9]In fact, *any* stochastic volatility approach discussed before, or the lognormal mixture diffusion model by Brigo and Mercurio (2002), could be used to model the implied volatility. For instance, Bandi et al. (2021) derive a close-form formula for short term options using Edgeworth expansions, and more recently, Guyon and Lekeufack (2023) highlight that volatility is path-dependent and present a SV factor model in which the diffusion term incorporates past returns.

On the other side of the spectrum, non-parametric models refrain from making explicit form assumptions about implied volatility, instead allowing the observed data to shape the surface. This class of models requires some form of smoothing, typically realised through kernel regression, (spline) interpolation, or local regression. Jackwerth (1999) and Bondarenko (2003) deliver an overview of the most prevalent non-parametric models. Particularly, Aït-Sahalia and Lo (2002) devise a non-parametric estimator for the Arrow-Debreu prices, or state-price density of option prices, providing a novel approach to estimate the implied volatility smile of options. Fengler (2009) introduces an arbitrage-free methodology for smoothing the implied S&P 500 volatility surface, while Gruber et al. (2020) implement a non-linear least square methodology on S&P 500 option prices from flexible multi-factor risk-neutral specification to extract option-implied factors. The natural extension is to perform a dimensional reduction. Both Skiadopoulos et al. (1999) and Alexander (2000) opened the door for a Principal Component Analysis (PCA) of the implied volatility. Alexander (2000) proposes a PCA on the ATM deviation of implied volatility smiles and presents the skew's principal representation. Cont and Da Fonseca (2002) conduct a multidimensional PCA (Karhunen-Loève decomposition) on the continuous implied volatility surface and fit the first three principal component time series to an AR(1)/Ornstein-Uhlenbeck process to discern changes in the IVS. Fengler and Härdle (2003) propose an alternative approach to modelling the dynamics of the implied volatility surface using the common PCA.

Recent publications in financial literature indicate a shift towards non-parametric Machine Learning (ML) techniques. Empirical evidence suggests that these applications exhibit superior performance in comparison to conventional statistical models across diverse financial areas, such as hedging (Buehler et al., 2019; Zhang et al., 2022), option pricing model calibration (De Spiegeleer et al., 2018; Horvath et al., 2021), the directional forecast of realised volatility (Kim and Won, 2018; D'Amato et al., 2022), or pricing volatility derivatives (Hoster et al., 2018).[10] For example, Cao et al. (2020) use a three-layer neural network to analyse S&P 500 volatility surface movements. The authors use daily options data between January 2010 and December 2017 to derive a relationship between expected volatility changes and index returns, moneyness and TTM. In essence, the authors fit a three- or four-factor model on the implied volatility surface using the NN. Although a comparative analysis of their fit and the resulting hedging parameter against other models is absent, they offer a primer into machine learning approaches for financial predicaments, highlighting the various benefits of different activation and loss functions. Others like Kim et al. (2021) implemented a more advanced transformer self-attention mechanism to generate a smooth equity index volatility surface under the SABR model. Similar to the deterministic models, no stochastic model has been applied to cryptocurrency implied volatilities, except for Matic et al. (2023).

---

[10]Dunis and Huang (2002) has already shown the advances of Neural Networks (NN) over statistical models. Both compare different ARCH-type models and benchmark them against NN predictions or simple IV and find that the outperforms statistical models in terms of RV prediction. Furthermore, Hamid and Habib (2014) find that NN outperform the statistical models in term of IV prediction of NIFTY options.

2.4.2 Forecasting

Modelling implied volatility not only benefits the pricing and hedging of options, but it also forms the bedrock for predicting future implied volatility, thus potentially providing a competitive edge for traders. Indeed, the forecast of implied volatility has held the academic fascination for several decades. The typical strategy for forecasting implied volatility involves initially identifying and determining its key drivers, modelling their dynamics, and subsequently using these predictions to extrapolate the surface. As early as when Harvey and Whaley (1992) propose a linear regression model, taking into consideration lagged values of the underlying, trading day dummy variables, lags of call and put IVS to account for autocorrelations, and interest rates to predict S&P 100 implied volatility smiles. However, the generated predictions failed to yield profit due to the eroding influence of transaction costs.

Cont and Da Fonseca (2002) adopt a non-parametric approach and predict both the S&P 500 and FSTE 100 volatility surfaces. Leveraging the Nadaraya-Watson estimator of Aït-Sahalia and Lo (2002) in conjunction with a Gaussian kernel, the authors create daily smooth volatility surfaces, subsequently performing a Karhunen-Loève decomposition on the surface's daily variations. This allows them to represent the surface as a function of the eigenmodes and eigenvectors originated from the principal component analysis. They show that the first three eigenvalues account for 98% (97.8%) of the S&P 500 (FTSE 100) variance and introduce a factor model to predict the implied surface. For this, the authors model the principal component through an Ornstein-Uhlenbeck process and translate the predicted components back to the surface. Their study underscores two pivotal points. First, implied volatilities exhibit robust positive autocorrelation and demonstrate mean reverting characteristics; and second, a substantial proportion of the daily log-variation in implied volatility's variance can be adequately explained by three principal components. By the same token (Beer and Fink, 2019) apply the same approach to various foreign-exchange surfaces. It's worth noting, however, that this PCA method yields optimal results when the correlation between parties is high, implying strong correlation between implied volatilities for diverse strike prices and maturities. Such correlation is not prevalent in cryptocurrencies. Furthermore, the PCA results are somewhat less transparent for traders, and the multi-step nature of the process offers little impetus for the implementation of these models in practical settings.

On the other hand, parametric models allow us to identify and capture risk factors, offering a reduced-dimension methodology where the IVS dynamics are governed solely by the estimated parameters. Gonçalves and Guidolin (2006) follow a two-step approach to predict the US equity surfaces. First, they use daily IVS of the S&P 500 over a four-year period starting from 1994 to fit a variation of the deterministic volatility function of Dumas et al. (1998) using a Ordinary Least Square (OLS) regression. Consequently, this technique yields a parameter time series which is then modelled through a vector autoregression. The forthcoming step involves forecasting the parameters for a one-step ahead timeframe and reconstituting the IVS based on these forecasts. The predicted IVS is then benchmarked against a random

walk and NGARCH(1,1), evaluated using a variety of statistical metrics. The researchers conclude that their VAR model forecast outperforms the random walk and NGARCH method in performance. Moreover, they provide evidence for the positive economic value of the forecast through numerous profitable volatility trading strategies. Expanding upon this methodology, Chalamandaris and Tsekrekos (2010) include the Nelson–Siegel term structure to generate implied volatility surfaces. Bernales and Guidolin (2014) extend the work of Gonçalves and Guidolin (2006) by considering not only indices but also single equity options. By analysing daily S&P 500 and 150 equity surface data from January 1996 to December 2006, they establish a time series of parameters. Here, they select Gonçalves and Guidolin (2006)'s identical polynomial volatility model and fit the surface using a generalized least squares (GLS) regression. Furthermore, they model the concurrent dynamics of equity and index implied volatility surfaces using two distinct VAR models and jointly through VAR with exogenous variables (VARX). Parallel to Gonçalves and Guidolin (2006), they backtest an ATM straddle and delta-hedge strategy and investigate their profitability. They conclude that, while their models surpasses the benchmark models such as random walk or NGARCH approach, transaction costs nullify the profit engendered from their forecasts. Other researchers, such as Kearney et al. (2018) experiment with different prediction approaches (functional time series) and show they can outperform the benchmark set by Gonçalves and Guidolin (2006) while Bedendo and Hodges (2009) use a Kalman filter to model the skew dynamics in order to assess whether this model is capable of forecasting the density of daily returns. All of these researchers report significant success in terms of forecasting a reasonable IVS and using this to develop profitable trading strategies. Despite the negation of profits due to transaction costs, it underscores the need to actively further pursue this strand of research and apply it to other asset classes.

Forecasting volatility using ML approaches is still in its infancy, and is focused exclusively on S&P 500 options. Chen and Zhang (2019) put forth a novel combination of an attention mechanism and LSTM model to anticipate one-step-ahead implied volatility surfaces, focusing on discrete grid points within a range of 20% above and below ATM, and maturities spanning from six months to two years, using raw IV time series as inputs. Medvedev and Wang (2022) use an LSTM and convolutional LSTM model to generate multivariate and multi-step-ahead forecasts of the S&P 500 IVS for one-, 30-, and 90-day horizons. The authors consider four fixed quarterly contracts and use a splicing technique to construct the volatility surface, i.e. contracts are stitched head to toe on expiration to form a continuous series. This, on the other hand, leads to an unbalanced panel with unequal time series observations for each contract. This gap is later filled by interpolating missing elements linearly. However, focusing only on the traded option prices is problematic in the presence of sparse or erroneous data. Furthermore, the extrapolation of such a major dataset may skew the final results. The authors benchmark their model against an AR, VAR, and VEC model, comparing their performance based on the model RMSE, and find that the ConvLSTM outperforms all other models. Furthermore, they test for statistical significance using the Diebold–Mariano test

statistics and introduce a profitable butterfly strategy. Lui et al. (2021) compare a traditional GARCH model with an Artificial Bee Colony NN for option trading. The authors simulate the underlying price with an MC simulation and compare different options strategies (straddle, butterfly, and calendar spreads). Bolch and Book (2022) fit an SVI model and create time series of the parameters. These parameters are then predicted using ConvLSTM, which then generate the IVS. The authors compare their prediction to a naive forecast and find that the ML approach outperforms the naive one on long-term forecasts on short to mid-range maturities. They conclude that the model is capable of predicting movements of the IV correctly. Similarly, Vrontos et al. (2021) investigate whether the direction of S&P 500 implied volatility and the VIX can be predicted. The authors perform a horse race of 20 different machine learning models and evaluate them based on the accuracy of their prediction and potential trading strategies. The authors derive profitable strategies from their predictions and encourage the use of various ML models (Naive Bayes, Ridge Deviance, Adaptive Boosting, Discriminant Analysis) for asset allocation. Lastly, it is noteworthy to mention that, as of this writing, the forecasting of bitcoin implied volatility remains an untapped research area within financial studies, although interest in predicting bitcoin returns is strongly present (Chen et al., 2022).

CHAPTER 3

BITCOIN OPTIONS DATA

At the time of writing, five major cryptocurrency exchanges, namely Deribit, CME, Delta, OKX, and Binance, offer options on bitcoin and other coins, as well as some tokens. Across all options exchanges, the aggregate average daily volume during March 2023 was above \$1.5bn. The volume traded on bitcoin options in particular, recently surged to all-time highs. Between January 2020 and January 2023 the average monthly trading volume increased more than 30-fold and the open interest increased more than ten-fold. To put this into perspective, the S&P 500 options market on the CBOE grew 'only' by about 17% between 2020 and 2023.[1]

The vast majority of trading is on the Deribit options exchange (>85%), which moved from Amsterdam to Panama and is now considering yet another move to Dubai, all in an effort to avoid adhering to international standards set by governmental agencies such as the CFTC or any other form of regulation protecting client interests.[2] Like many other unregulated crypto derivatives exchanges that are typically registered in off-shore tax havens, Deribit's trading platforms are open 24/7 and there is little or no compliance with know-your-customer protocols. Deribit options have (bi-)daily, (bi-)weekly, (bi-)monthly and quarterly expiry, up to nine or twelve months and are European-style and crypto/cash-settled.[3] The underlying is the Deribit Bitcoin Index (DBTC) which is an equally-weighted average of the latest bitcoin price on four to eleven exchange: First, the exchange calculates the median of the available prices and excludes those 0.5% above and below the median. The remaining prices are equally

---

[1]See CBOE Historical Options Data for trading volumes on SPX options on the CBOE.

[2]See CCData Research Report or The Block Option for cross-exchange trading data and visualisation.

[3]A new option expiry is introduced every Thursday immediately prior to the expiration with the following exceptions: A weekly option will not be introduces if there already exists a monthly expiry. This monthly option will now be the weekly expiry. A monthly option will not be introduces if there already exists a quarterly expiry. This quarterly option will now be the monthly expiry. Note that the minimum order size is 0.1 options with a minimum tick size of 0.0005XBT and a block trade contains a minimum of 25 contracts. The standard call option's strike range goes from 0.1 to 0.9 delta and -0.1 to -0.9 for put options, respectively. See Deribit Option Specification accessed on 01 May 2023.

weighted and form the index. Currently, the exchanges include Bitstamp, Coinbase Pro, Kraken and LMAX Digital and the index is updated every second.[4] There are more option maturities than futures maturities. Nevertheless, Deribit provides a futures value for the missing maturities by their so called 'synthetic-futures'. These artificial, not tradable futures serve as underlying for options although the Deribit website is rather contradicting. The exchange present the (synthetic) futures as underlying when calculating the option's implied volatility and Greeks, but the options specification documents state that the underlying is the Deribit bitcoin index, which does not imply that the (possibly synthetic) futures contract is the underlying. As of now, there is no publicly available methodology on how these synthetic products are priced. The option strike ranges vary from 40% to 180% of the current underlying price for shorter maturities and up to 1100% over the current bitcoin price for maturities more than nine months, while the increments between strikes varies from $250 for short-term maturities, to $10,000 for longer dated options.

The sheer size of trading volume on Deribit makes it the most attractive exchange to consider for any type of cryptocurrency option research. Even though the CME (and a few other exchanges) list bitcoin options, only 10%-15% of the total volume traded on bitcoin options has ever been attributable to these exchanges. Deribit alone accounts for over 85% of bitcoin options trading volume.[5] One reason might be that Deribit operates 24/7, whereas the CME closes on weekends and holidays. Another may be that Deribit options are margined and settled in bitcoin, even though the underlying is the US dollar value of the BTC index.[6] Their attraction lies in being able to trade a derivative on a fiat-crypto cross without using fiat currency for collateral in the margin account, or for settlement of the contract. Another explanation might be that investor's money on Deribit has been safe from hacks, as there has been no major security breach affecting trader's money. Additionally, the exchange offers new products frequently, and is among the oldest in this space with almost 100% uptime.

This chapter presents the technical setup which allows us to create our proprietary data and explains common data filtering and interpolation techniques applied to the bitcoin options data which form the backbone of this thesis. Furthermore, it provides a descriptive statistic of the trading volumes and open interest of derivatives traded on the Deribit exchange and motivates our further work. We divide the sample into different maturity and strike classes and highlight trading clusters for bitcoin option.

---

[4]This methodology has changed recently. Prior to February 2023, the index was an equally weighted index of prices on eleven exchanges, where the highest and lowest price is excluded and the remaining nine are used to calculate the index. The used exchanges included Binance, Bitfinex, Bitstamp, Bittrex, Coinbase Pro, Gemini, Huobi Global, Itbit, Kraken, LMAX Digital and OKX. For a detailed specification, see Deribit Index Calculation accessed on 01 May 2023.

[5]Second comes the CME (5%), then OKEx (2.5%) and Bit.com, see The Block for more details.

[6]To obtain the maturity payoff one needs to calculate the difference between the BTC value in dollars and the option strike and convert the result to bitcoin using the BTC index value at maturity. To calculate the final payoff, Deribit uses the 30-minute average of of the BTC Index prior to expiry as settlement value, see the official Deribit Options Specification.Note that the Deribit bitcoin options market is not complete. The index itself is not tradable and requires costly replication and frequent rebalancing. The lack of information on the precise calculation of the settlement value results in an incomplete market for traders, see Chapter 4.

3.1 Data Sourcing & Manipulation

All Deribit cryptocurrency derivatives market data is publicly available, free of cost and in real-time accessible through their Application Programming Interface (API). The exchange offers a vast variety of different information, including but not limited to, the current order book for all products offered, the funding rate, and the index level, see Deribit Public API for all publicly available data.

Sourcing

Despite the broad range of data accessible, our study primarily concentrates on market data, specifically details related to option trades, associated prices, trade sizes, and underlying assets. For the extraction of this precise set of data, we utilise a pair of commands that are sequentially executed within a matter of milliseconds:[7]

- `/public/get_book_summary_by_currency`:

  This command returns an `json` file containing the current orderbook summary of the designated cryptocurrency, in this instance, bitcoin. It imparts comprehensive data related to each accessible option and futures derivative, which includes the instrument name and corresponding underlying, current best ask-, bid- and mid-price, timestamp of creation, and the prevailing funding rate. It further provides details about the maximum and minimum transaction prices during the preceding 24 hours. Additionally, it offers information on the last transaction price, the number of contracts traded and the corresponding notional in the past day, along with the current open interest.

- `/public/get_index`:

  This returns the underlying Deribit Bitcoin Index value in USD.

The integration of data derived from these two commands captures the entire state of the bitcoin options market on Deribit. It is important to note that these commands only provide instantaneous snapshots of the market state at the current moment, with no possibility to fetch historical trades data. To accomplish this, we set up a dedicated server and installed a `crontab` job to automate the script execution. This approach has allowed us to create our own proprietary dataset of historical bitcoin options data, spanning from January 2020 to January 2023, with a minute-by-minute frequency. In the rare instances of server malfunctions, we have retrieved missing data from reliable external sources, such as CCData.

---

[7]The information is based in the API Version 2 and are for till January 2023. Any update and/or adjustment with the commands or information output after January 2023 is not covered in this thesis.

Filtering

The bitcoin options market is not a high-frequency market, with trading and liquidity primarily clustered for short-term ATM options, as will be confirmed in subsequent analysis. This low-activity environment can produce misleading market impressions, resulting from outdated, or stale, prices for options that are not actively traded or have unusually large bid-ask spreads. In an effort to mitigate these effects, we use the mid-price of options, in line with the approach proposed by Jackwerth and Rubinstein (1996). While we understand that the true value of options is best reflected by the prices at which they are traded, the continuous updates of bid-ask prices by market makers can serve as a fair representation of an option's authentic value. To enhance the robustness of our analysis, we filter out options that have no trading volume in the past 24 hours, as well as those with a percentage bid-ask spread (BAS) greater than 25%. We calculate the percentage bid-ask spread by subtracting the bid-from the ask-price, then dividing the result by the mid-price. This filtration is employed to exclude options that are not offered by credible market participants. Given that this market is still developing, it is common to encounter 'joke spreads' – for example, a deep in-the-money (ITM) option with a bid-price reflecting a 490% implied volatility and an ask price of 80%. Our findings reveal that the average percentage bid-ask spread for ATM (OTM) options is roughly 4% (10%), rising to 5% (15%) for options with extended maturities.

Moneyness

The implied volatility of bitcoin, with its sudden, wide swings in value, presents a challenge when comparing specific strike levels over an extended time horizon. To address this concern, and to facilitate clearer interpretation of the outcomes, we transition from a strike to a moneyness metric. We define moneyness m as:

$$m = \frac{K}{S},$$

where K is the strike level of the option and S the underlying value in accordance with the approach outlined by Alexander et al. (2012). A 'delta-metric' could potentially obscure the results for shorter maturities, as it could consider a narrow strike range in absolute terms. This is due to the nature of bitcoin options, which are characterised by relatively short maturities and high volatility. This implies that, in translating strikes into deltas, we could find a very low delta, e.g. 15 or 20, already after three or four strikes away from ATM. Consequently, it would result in two or three different grid points falling between adjacent strikes. Specifically, if we were to interpolate, we might be interpolating different deltas derived from the same two options. This could produce spurious dependence which would distort results.

Interpolation

In the event that an option is filtered out, we create a synthetic option price through interpolating the implied volatility rather than option prices themselves, following Galai (1979) who posits that direct interpolation of the option prices could yield non-investable returns. For this, we calculate the implied volatilities of all available market option prices by inverting the Black and Scholes (1973) option price formula, i.e. we approximate the implied volatility numerically. However, we encounter instances where the maturities of the options do not align perfectly with those of the underlying futures, resulting in a discrepancy in the number of options and futures maturities.[8] Moreover, liquidity is a major consideration with finite-maturity futures contracts. Stale futures prices could introduce inaccuracies in the option's price, and by extension, its Greeks. Given these potential pitfalls, we favour the use of the Put-Call parity (PCP) relationship to infer the correct futures price, instead of using its market price. Naturally, the PCP values will vary for each strike level. Due to the general focus on ATM options in trading, it is often challenging to identify an ITM/OTM strike level where both calls and puts are actively traded. Thus, we employ the PCP derived from ATM options and use this value for the entire option chain. We interpolate the ATM PCP values of two neighbouring maturities and use this to compute the synthetic fixed-maturity underlying of the option and calculate the implied volatility.[9] Due to the clusters of trading and liquidity, which will be discussed in the succeeding section, we only consider OTM options when constructing the raw grid of the implied volatility surface.

The landscape of interpolation techniques offers a plethora of options that vary significantly in complexity, ranging from linear interpolation over the implied variance, to higher-degree polynomials or kernel regressions. Similar to other areas of financial literature, there is not one single correct interpolation technique, but many wrong ones. We follow the empirical literature on option pricing and hedging and use the implied volatility shape-preserving Piecewise Cubic Hermite Interpolation Polynomials (PCHIP) as advocated by Madan and Milne (1994), Bliss and Panigirtzoglou (2002) or Lim and Ting (2013) among others. This method boasts several advantageous attributes: it maintains the shape and monotonicity of the implied volatility data points while providing a smooth surface free of oscillation between two points. Moreover, it proves to be highly computationally efficient. The polynomial splines are defined as follows:

---

[8]For example, an option might expire on 07 January 2023 but there exist no fixed-maturity futures with same maturity. For these instances, Deribit provides 'synthetic' futures prices with same maturity as the options that are used to calculate the options IV and Greeks. However, the exchange does not provide a methodology on how they calculate the synthetic options prices. However, we believe that it is a simple interpolation between traded futures prices with straddling maturities.

[9]In theory, due to the PCP, both put and call option should have the same implied volatility for an identical set of inputs. Given the strongly tilted interest towards trading OTM options compared to ATM or ITM, we decide to consider the implied volatility for OTM options as these reflect greater stability and information content. For the ATM case, we focus on call options. We have compared both call and put options for ATM strikes with different maturities and find no significant difference between their implied volatilities. Hence, we use the call implied volatility as proxy for further calculation.

Definition 1 (Cubic Hermite Spline). Let $f : [a, b] \to \mathbb{R}$ be an unknown function with known values for $((x_1, f(x_1), (x_2, f(x_2), ..., (x_n, f(x_n))$ and their respective slopes $f'(x_1), ..., f'(x_n)$ where $a = x_1 < x_2 < ... < x_n = b$. We define the Hermite basis functions as:

$$h_{00}(x) = (1 + 2x)(1 - x)^2, \qquad\qquad h_{01} = x^2(3 - 2x)$$
$$h_{10}(x) = x(1 - x)^2, \qquad\qquad h_{11} = x^2(x - 1),$$

and define an additional function $z_i(x) = \frac{x - x_i}{x_{i+1} - x_i}$. The cubic Hermite spline function $\hat{f}$ is then given by:

$$\hat{f}(x) = \begin{cases} f(x) & \text{for } x \in \{x_1, ..., x_n\} \\ h_{00}\left(z_i(x)\right) f(x_i) & \text{for } x_i < x < x_{i+1} \\ \quad + h_{01}\left(z_i(x)\right) f(x_{i+1}) \\ \quad + h_{10}\left(z_i(x)\right) (x_{i+1} - x_i) f'(x_i) \\ \quad + h_{11}\left(z_i(x)\right) (x_{i+1} - x_i) f'(x_{i+1}) \end{cases}$$

We use the `interpolate` package of `SciPy` to estimate the interpolation function according to the algorithm proposed by Fritsch and Butland (1984) which results in a piecewise cubic Hermite interpolating polynomial that is continuously differentiable and preserves the original shape of the bitcoin implied volatility smile.

No-Arbitrage Constraints

The foremost consideration, regardless of the interpolation technique employed, is the assurance that the interpolated volatility surface implies arbitrage-free option prices. We investigate the interpolated option prices, ensuring that the essential no-arbitrage constraints, as outlined by Carr and Madan (2005) and Fengler (2009), are satisfied for both mid-prices and interpolated values. We denote the (interpolated) implied volatility at time t as $\sigma := \sigma_t(K, T|S)$ for any maturity T, and the corresponding BS call option price at time t as $C_t(S, K, \tau, \sigma)$, where τ denotes the option's time to maturity, given in year. For the sake of clarity, we avoid any sub- or superscript for the interpolated implied volatility, nor do we define it as a function of strike and maturity although we are aware that each (interpolated) implied volatility is specific to an option.[10] Note that the interest rate r is deliberately omitted given our assumption of a zero interest rate. During the period of 2020 to 2023, interest rates were historically low. Given the short maturities of bitcoin options, these rates become essentially negligible. Similarly, we define the BS put option price as $P_t(S, K, \tau, \sigma)$ and confirm that the (interpolated) option prices are positive and comply with the following conditions:

---

[10]Note that the PCHIP interpolation is shape-preserving, i.e. the market implied volatility equals its 'interpolated' counterpart for a traded options which have not been filtered out.

- **Monotonicity:** Call (put) option prices are increasing with maturity and decreasing (increasing) in strike levels:

$$\frac{\partial C_t(S, K, \tau, \sigma)}{\partial \tau} \geq 0 \qquad \text{and} \qquad \frac{\partial P_t(S, K, \tau, \sigma)}{\partial \tau} \geq 0,$$

and

$$\frac{\partial C_t(S, K, \tau, \sigma)}{\partial K} \leq 0 \qquad \text{and} \qquad \frac{\partial P_t(S, K, \tau, \sigma)}{\partial K} \geq 0.$$

- **Convexity:** Call and put option prices are convex functions with respect to the strike for all available τ:

$$\frac{\partial^2 C_t(S, K, \tau, \sigma)}{\partial K^2} \geq 0 \qquad \text{and} \qquad \frac{\partial^2 P_t(S, K, \tau, \sigma)}{\partial K^2} \geq 0.$$

- **Total Variance:** The total variance $\nu^2(S, K, \tau) = \sigma^2 \tau$ is increasing with maturity:

$$\frac{\partial \nu^2(S, K, \tau)}{\partial \tau} > 0.$$

- **Boundaries:** The prices of call and put option are bounded up- and downwards. For $\forall K, \tau > 0$ it holds:

$$(S_t - K)^+ \leq C_t(S, K, \tau, \sigma) \leq S_t,$$
$$(K - S_t)^+ \leq P_t(S, K, \tau, \sigma) \leq K.$$

3.2 Trading Data

Figure 3.1 depicts the daily evolution of the settlement price of bitcoin options on Deribit, i.e. the DBTC index, at midnight UTC and the total traded volumes (as notional amounts, in \$bn) of all options, perpetuals and the fixed-maturity futures on Deribit over the previous 24 hours between January 2020 to January 2023. During 2020, from a level of around \$7000, the DBTC index rose relatively slowly until its first major bull run started in November 2020, and the index value reached almost \$28,000 by the end of 2020. During 2021 the DBTC index more than doubled between January ($\sim$ \$28,000) and mid April 2021 ($\sim$ \$59,000) then fell almost 50% until mid July (\$30,000). Its all-time high on 8 November 2021 was around \$69,000 before it started to decline continuously. The first four months of 2022 show a slight and range-bounded increase before the bitcoin price crashed in May 2022 due to the TerraUSD (UST) de-pegging from the dollar. It stayed fairly stable afterwards around \$20,000 and crashed yet again in November 2022 due to the FTX fallout. The second plot of Figure 3.1 shows that total 24-hour trading volume over all Deribit options was relatively low during 2020, barely exceeding 500 million USD daily volume, and really took over by the end of 2020. However, during 2021 there were pronounced periods of volatile or directional markets

where daily options volumes of up to 3 billion USD became commonplace. The number of different options contracts traded also almost doubled, from 4.3 million in January 2021 to 6.2 million by the end of the year. Throughout 2022, option trading volume remained stable with around $750m daily notional. Given that the underlying price declined from $45,000 to $18,000, it indicates that the number of traded contracts rose during the last year. In fact, with over 7.2 million traded contracts in 2022, the exchange recorded yet another 16% increase in traded contracts year-to-year. The two plots underneath depict the daily trading volume for the perpetual (orange) and fixed-maturity futures (yellow), which saw the most active trading during 2021, especially during the first half of the year, than in any other period. Interestingly, the second half of 2021 showed weaker growth in trading on the perpetual than for options. The latter may have been driven by the introduction of many new types of contracts in late 2020 and early 2021, which traders gradually adopted for gamma and vega hedging. By the second half of 2021, this could have diminished the pressure for extremely active dynamic delta-hedging. In fact, the volumes traded on futures contracts also fell off – even more so than for the perpetual – during the last six months of 2021 and continued throughout 2022. In any case, finding that trading patterns during 2020 and 2022 were so different from those in 2021 motivates a data split for any empirical model comparison.[11]

Next, we examine some empirical characteristics of the bitcoin perpetual contract and compare these with fixed-maturity futures. Since the settlement price for bitcoin options is not for a tradable contract, any trader needs to consider either the futures or the perpetual if they want to hedge their position. In this case, the effectiveness of hedging an option with a futures contract depends, among other things, on the variability of the basis. To illustrate this variability, Figure 3.2 depicts the difference between the market price of the futures (or the perpetual contract) and the DBTC index, divided by the DBTC index, from January 2020 to January 2023. This percentage basis is given in basis points (bps), on the left-hand scale for the three synthetic fixed-maturity futures, and on the right-hand scale for the basis relative to the perpetual. As a result of the funding rate mechanism, the basis risk of perpetual futures is very low – most of the time it is less than ± 10 bps. But it is also highly variable – for instance, during the crash in March 2020 the perpetual basis reached almost -150 bps. The very tiny basis risk of perpetuals indicates that they could offer better delta-hedging instruments than the calendar futures of the same maturity as the option. Unlike the perpetual basis, the bases for the fixed-maturity futures are almost always positive. For 10-day futures the basis can be as high as 100 bps, and for longer term futures the basis can even reach 450 bps. Also notice from this figure that the futures curve at 10, 20 and 30 days maturity is normally in contango – in fact, the ordering $F_{30} > F_{20} > F_{10}$ is present for 87% of the time, i.e. 953 days of the 1095-day sample and only in backwardation during March/April 2020 (COVID crash and

---

[11]Options outnumber futures by a factor of two. Deribit has the first mover advantage on options while Binance and others have that on spot and futures. Given the new influx of capital and participants, the question of 'where should we trade' becomes easy to answer as Binance has all the liquidity and orderbook depth for futures, Coinbase and Kraken for spot and Deribit for options.

Figure 3.1: Bitcoin Price Evolution and Daily Derivatives Notional Trading Volumes

The upper plot illustrates daily BTC index prices at 00:00 UTC over a three-year sample period starting on 1 January 2020 (top, black plot); next the corresponding 24-hour total notional trading volume on all Deribit options (second from top, blue plot); and the daily notional trading volume on the perpetual (third from top, orange plot); and finally, the daily notional trading volume of fixed-maturity futures (bottom, yellow plot). Values are given in $10,000 for the BTC index and $bn for options and futures trading volumes.
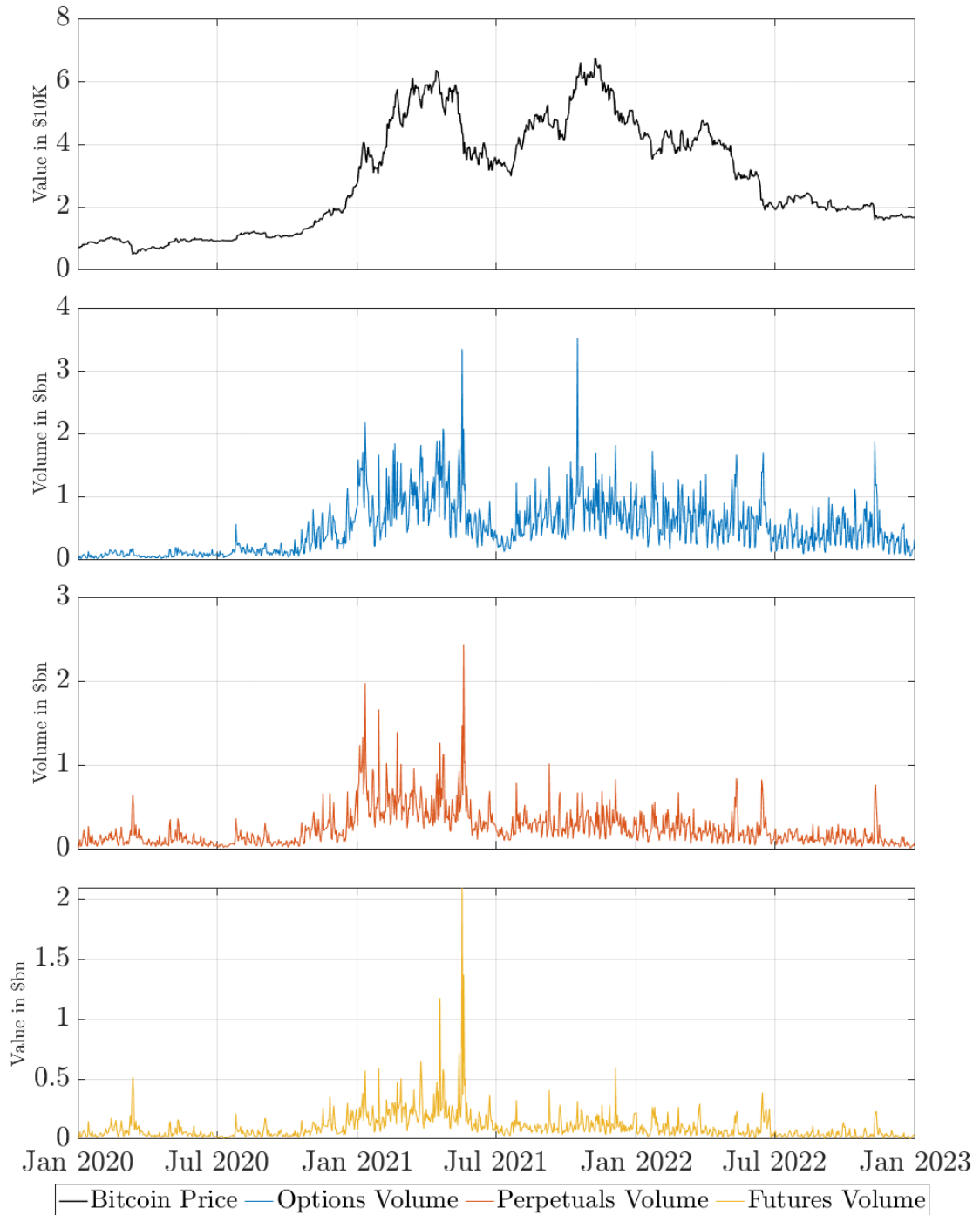
its aftermath), June 2021 (end of a bull run), May 2022 (UST de-peg) and November 2022 (FTX collapse). Surprisingly, there is virtually no difference between perpetuals and standard futures after February 2022, i.e. the difference between 10- and 30-day basis is negligible.

Figure 3.2: Perpetual and Futures Basis

The futures price minus the BTC index, divided by the BTC index, in basis points. The right-hand scale measures the percentage basis for the perpetual futures (black) and the left-hand scale measures the percentage basis for the futures with constant maturities of 10, 20 and 30 days (in blue, green and red, respectively). The sample covers three-years starting in January 2020 and daily snapshots are taken at midnight UTC. Values are expressed in absolute bps.
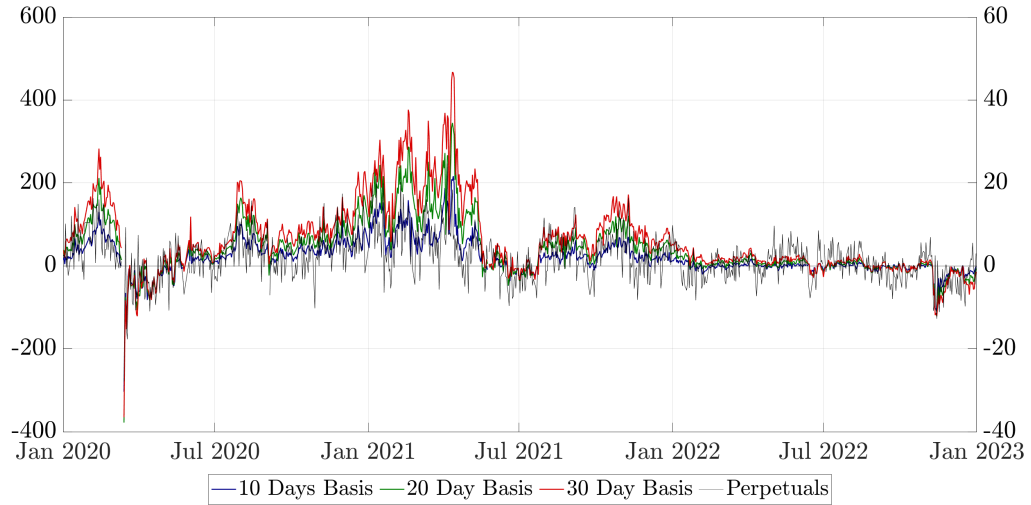
Alexander et al. (2022) document many differences between the bitcoin and S&P 500 options markets. One of the main differences is the proportion of short-, medium- and long-dated options that are traded. A one-month option on the S&P 500 index is considered relatively short term, because the majority of trading occurs between the one-month and the three-month expiries. However, a bitcoin option with one month to expiry falls into the longer-term category due to bitcoin's high volatility. Though monthly or quarterly equity options are most frequently traded, these maturities display comparatively lower volumes for bitcoin. To see this, Figure 3.3 depicts the proportions of daily traded contracts (top plot) and the open interest (bottom plot) on Deribit based on their time to maturity from January 2020 to January 2023. On the right-hand scale, the solid black line represents the total number of traded contracts traded over all expiries. For the sake of clarity, we present these data using their weekly average on a rolling window. The seasonal pattern in the proportion of traded contracts of short-term (up to two-weeks) options is a result of the issuing schedule policy, whereby options with one week (and/or two weeks) to expiry are issued unless there is a standard monthly or quarterly option expiring with the week (or two weeks). On the left-hand scale we present the proportion of short-term (up to two weeks), mid-term (between two weeks and a month) and long-term (more than a month) maturities. For this, we accumulate

all daily traded contracts within each expiry class and present it as percentage over all traded contracts, again using a weekly average on a rolling window for the sake of clarity. Apart from this seasonal pattern, during the entire three-year period only around 15%–20% of trading is on options with expiry dates longer than one month. Even though the number of traded contracts has continually risen over our sample, the proportion of contracts with more than one month to expiry has remained relatively constant, as has the proportion of short term options of up to two weeks. In fact, around 60% of all traded contracts is on these very short-term maturities. Another 20%–25% corresponds to mid-term options with maturities between two weeks and one month. Given that the options with maturity of up to one month constitute 80%–85% of all trading volume on Deribit, we decided to focus on these options.

We observe an entirely different picture with the open interest. Similar to the traded contracts, we present the daily open interest as a weekly rolling average. Approximately half of the open interest rests in options with expiry dates exceeding a month. Notably, the highest level of open interest is concentrated in quarterly options, despite their less frequent trading activity relative to short and mid-term maturities. Investigating further into trading patterns, we find that short- and mid-term maturities are primarily traded in small quantities (fewer than one contract) on an almost hourly basis. In contrast, longer-term maturities are traded in larger volumes, with some trades reaching up to 25 contracts (block trades) per transaction. This trading pattern implies that long-term maturities are likely used for hedging purposes, as opposed to speculative trading. Mid-term options represent 15%–20% of the total outstanding contracts, while short-term maturities account for approximately 30%–35%. Here again, we observe a seasonal pattern due to the transition of quarterly maturities from long to mid-term as they approach their expiration dates. This pattern suggests a roll over of positions by traders as these quarterly options reach their expiry.

Tables 3.1 and 3.2 complete the options data chapter. Motivated by the clear distinction of traded volumes and open interest across different maturities, as seen in Figure 3.3, Table 3.1 shows the total number of traded call and put contracts ('Trades'), the average number of available strikes ('Strikes') and the average end of day open interest ('OI') for different maturities and moneyness subcategories from January 2020 to January 2023. We focus only on OTM options, i.e. $m < 1$ for puts and $m \geq 1$ for calls as these show much more trading activity and liquidity. Cumulatively, more than 11.8 million OTM contracts have been traded across all maturities, with a mean price of \$1308 and a corresponding implied volatility of 86.5%. Every day, traders had access on average to over 156 distinct strike levels, but preferred to trade more call contracts (6.3 million) than put options (5.5 million).

The short-term derivatives emerge as the most traded assets in this context. Both ATM call and put options account for over 1.8 million contracts traded each within the previous three years. This moneyness range also provides the most tradable strike levels, i.e. approximately twelve options on either side. Indeed, even when all other moneyness ranges for this maturity class are aggregated, they do not match the volume of ATM trading. Simultane-

Figure 3.3: Maturities of Traded Options and Open Interest

Proportions (left-hand scale) of total trading volume (upper plot) and the open interest (lower plot) in short-term options (up to two weeks, dark grey), mid-term options (between two weeks and one month, mid grey) and long-term options (longer than one month, light grey). The black line (right-hand scale) depicts the total number of traded options contracts. All series are a weekly rolling average of daily data from January 2020 to January 2023.
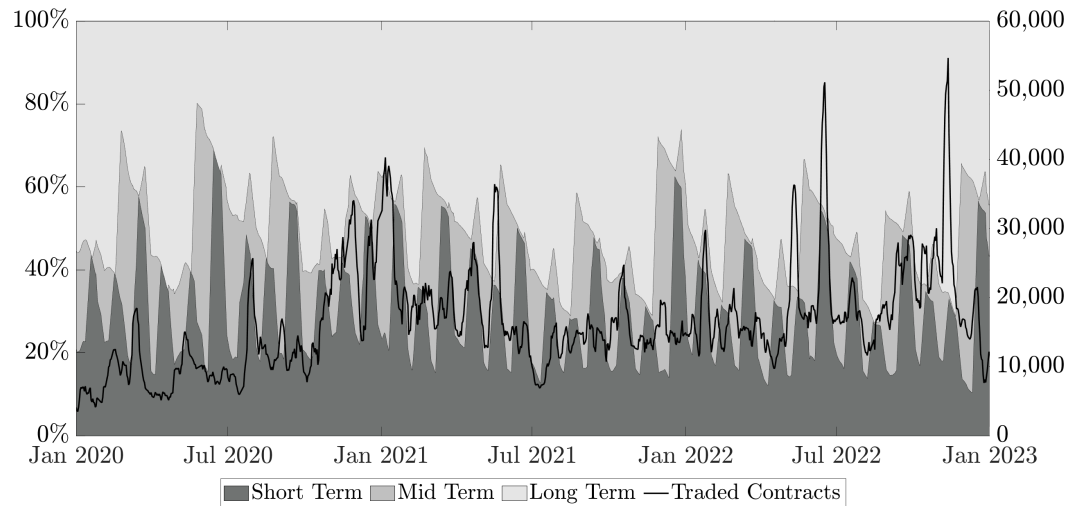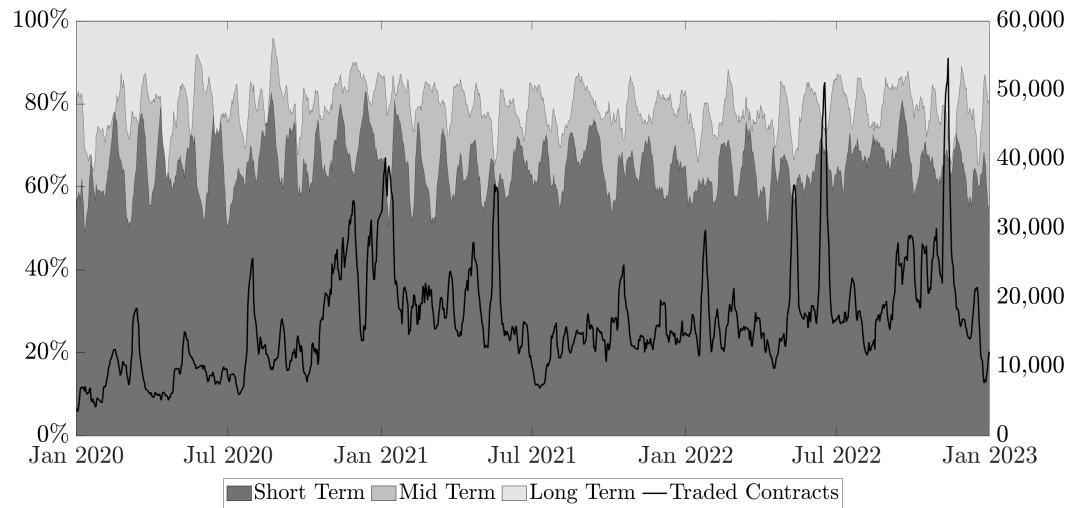
ously, the ATM calls and puts exhibit the smallest average daily open interest, suggesting high-frequency and speculative day trading, presumably dominated by algorithms or algorithmic trader. As we move further OTM within the short-term maturity bracket, we observe a decline in trading activity and available option strikes. However, there is a concurrent increase in open interest. A plausible explanation for this trend is that traders might be speculating with short-term OTM options, willing to hold onto them until maturity given their appealing risk-return profile. This hypothesis is supported by the heightened trading volume and open interest for deep OTM calls ($m > 1.7$). Recall that the short-term maturity category considers options expiring within the next two weeks. Thus, it seems unlikely that institutional traders would retain options with such short maturity and strikes situated 70% above the current underlying level. Rather, this trend might indicate retail trading activities, specifically those speculating on price jumps while holding options expiring in 0, 1, or 2 trading days. An alternative, albeit less probable explanation, could be related to liquidity constraints, potentially 'forcing' traders to hold their option contracts until maturity rather than closing the position due to a wide spread which would nullify any profit.

The findings for mid-term maturities deviate from those of short-term maturities modestly. The bulk of the trading volume remains concentrated around the ATM level, reaching its peak slightly above the immediate ATM, notably within a range of 10%-20% above and below the current underlying price. We continue to observe the decreasing trend in trades and available strikes as we progress further out of the money, albeit at a less pronounced rate, with the sole exception being the deep OTM call options. The trade volumes for mid-term options are approximately a quarter of those for the short-term options; at the ATM level, this ratio further diminishes to a fifth. Nonetheless, we find a pivot point for strike levels 30% (40%) above (below) the current underlying price, where mid-term manifest a higher trading volume. In a similar vein, the average number of tradable strikes and open interest for ATM are substantially lower, yet they pivot as we move further away from ATM. Across the entire options chain, mid-term options exhibit an average of 25 tradable strike levels, which is only half of what their short-term counterparts offer, signaling an increase in strike increments. The relatively balanced number of contracts traded around the underlying price level, i.e. ±20%, may serve as an indicator for potential options trading strategies, such as strangles or butterfly spreads, pointing towards institutional influence.

The traded contracts distribution for short- and mid-term options resembles a normal distribution in shape with a fat right tail, while the long-term options display a fairly uniform distribution. Recall that long-term options consider maturities from one month to a year, yet the trading volume associated with these options amounts to only half of that witnessed for options with maturities spanning from one day to two weeks. These longer-term options are notably favoured for large-volume trades (block trades), often comprising 25 contracts within a single transaction. Interestingly, there is an excess on the deep OTM calls for both, trades and open interest. Roughly one in every five long-term call options traded is set at a strike

price twice that of the current underlying. In fact, the far deep OTM call options ($m > 2$) display a higher trading volume than any near-the-money subclass. Despite this moneyness range having almost three times more available strike prices, it signifies a highly speculative trading pattern. Nonetheless, for empirical research related to trading activity, long-term maturities prove to be less well suited. Options with maturities greater than one month have exhibited too many stale prices, even at the hourly frequency, to be useful in any empirical analysis. Additionally, large-volume trades distort the representation of trading volume and frequency. Consequently, we select 10-, 20-, and 30-day constant maturity options for our trading research, each representing a proxy for the three main expiry classes documented.

By the same token, Tables 3.2 considers the same maturity and moneyness subgroups but presents the average implied volatility ('iVol') paid for the options in each subcategory, the average bitcoin price paid ('BTC') and the US dollar ('USD') equivalent. Constructing an implied volatility smile using the average traded implied volatilities for short-term options yields a fairly symmetric shape, although it demonstrates a slight excess on the OTM puts and lowest value for ATM levels. Particularly for the ATM range, i.e. 40-60 Delta we find a symmetric implied volatility. In contrast, the medium-range maturities' implied smile follows more of a hockey-stick shaped skew with marginal variation for ATM and slight increase at OTM call strike levels. On the other hand, the long-term maturities illustrate a relatively flat implied volatility curve with little difference between ATM and deep OTM strikes. Interestingly, the term structure on average resembles a fearful market condition. Typically, the term structure for equity indices shows an increasing ATM implied volatility with respect to maturity; this pattern is inverted only during market crashes. Such crash-like behaviour is observed on average for bitcoin, where the ATM short-term implied volatility surpasses those of longer-dated options. However, a more comprehensive examination of the implied volatility behaviour will be presented in Chapter 5. Naturally, the high volatility of bitcoin results in higher option prices, barely affordable for retail traders. For instance, an ATM long term option costs about $4700, even short maturities ask north of $550. Given this trading data, we conclude that the average bitcoin price (USD price divided by BTC price) was around $31,000.

Table 3.1: Summary Statistic: Trades, Strikes and Open Interest

Summary statistic of traded bitcoin option contracts on Deribit. We consider OTM calls and puts with varying moneyness levels and divide in short-term options (up to two weeks), mid-term options (between two weeks and one month) and long-term options (longer than one month) as well as 'Total' reflecting all maturities/moneyness levels. The sample covers trades between January 2020 to January 2023 but does not consider ITM options. The column 'Trades' presents the number of traded contracts in the respective moneyness and maturity category. The 'Strike' column provides the average number of tradable strikes in the respective moneyness and maturity category. Similar, the 'OI' column shows the average number of end-of-day open interest in each moneyness and maturity category.

	Moneyness	Short-Term			Mid-Term			Long-Term			Total		
		Trades	Strikes	OI	Trades	Strikes	OI	Trades	Strikes	OI	Trades	Strikes	OI
	$m < 0.5$	8,983	0.13	846	21,311	0.79	884	145,384	8.03	638	175,579	8.96	663
	$0.5 \leq m < 0.6$	34,072	0.54	821	51,928	1.14	726	114,605	4.08	662	200,606	5.76	690
	$0.6 \leq m < 0.7$	120,575	1.55	711	85,863	1.51	857	187,473	4.63	652	393,876	7.96	704
Put	$0.7 \leq m < 0.8$	322,031	3.88	585	190,503	2.13	820	254,433	5.22	665	766,967	11,23	667
	$0.8 \leq m < 0.9$	884,592	7.07	591	345,633	2.77	816	326,138	5.74	605	1,556,363	15.58	636
	$0.9 \leq m < 1$	1,826,346	12.3	401	326,814	3.05	573	248,470	5.29	461	2,401,631	20.63	442
	$1 \leq m < 1.1$	1,880,291	11.8	456	344,449	2.95	679	250,958	5.28	588	2,475,698	20.03	523
	$1.1 \leq m < 1.2$	752,401	6.32	590	357,160	2.61	841	256,538	4.96	652	1,366,099	13.88	659
	$1.2 \leq m < 1.3$	241,781	3.48	567	214,005	1.99	939	262,422	4.45	737	718,208	9.92	718
Call	$1.3 \leq m < 1.4$	94,026	1.73	591	116,778	1.46	1052	224,446	3.65	846	435,251	6.84	825
	$1.4 \leq m < 1.5$	43,488	0.85	792	72,178	1.46	961	188,001	3.37	889	303,668	5.35	889
	$1.5 \leq m < 1.6$	23,828	0.45	769	37,893	0.9	996	146,096	2.96	947	207,818	4.31	938
	$1.6 \leq m < 1.7$	12,709	0.21	1,040	33,242	0.62	1,016	115,117	2.62	911	161,069	3.44	938
	$1.7 \leq m < 2$	18,290	0.22	1,106	43,213	0.82	1,362	223,228	6.23	996	284,731	7.26	1,041
	$m \geq 2$	6,415	0,07	1,515	21,909	10.55	1,309	373,696	14.7	966	402,013	15.32	981
Total		6,269,833	50.59	528	2,262,876	24.42	844	3,316,972	81.21	758	11,849,681	156	697

Table 3.2: Summary Statistic: Prices

Summary statistic of traded bitcoin option prices on Deribit. We consider OTM calls and puts with varying moneyness levels and divide in short-term options (up to two weeks), mid-term options (between two weeks and one month) and long-term options (longer than one month) as well as 'Total' reflecting all maturities/moneyness levels. The sample covers trades between January 2020 to January 2023 but does not consider ITM options. The column 'iVol' presents the average implied volatility paid for the traded contracts in the respective moneyness and maturity category. The 'BTC' column provides the average price paid in bitcoin for the traded contracts in the respective moneyness and maturity category. Similar, the 'USD' column shows the average USD value of a traded bitcoin option in each moneyness and maturity category.

	Moneyness	Short-Term			Mid-Term			Long-Term			Total		
		iVol	BTC	USD	iVol	BTC	USD	iVol	BTC	USD	iVol	BTC	USD
Put	$m < 0.5$	202.46%	0.0014	39.5	156.28%	0.00195	64.15	106.14%	0.01168	413.01	112%	0.01067	376.58
	$0.5 \leq m < 0.6$	164.65%	0.0019	59.57	126%	0.003284	100.72	91.02%	0.02659	873.82	104.82%	0.01968	645.04
	$0.6 \leq m < 0.7$	137.79%	0.00266	83.42	107.09%	0.00561	179.46	83.45%	0.04491	1376.78	99.07%	0.02865	880.15
	$0.7 \leq m < 0.8$	114.26%	0.00405	118.35	91.42%	0.01044	338.69	78%	0.06454	1835.88	93.07%	0.03338	958.66
	$0.8 \leq m < 0.9$	96.73%	0.00723	219.55	78.97%	0.02133	714.73	74.05%	0.09429	2852.82	85.21%	0.04183	1278.44
	$0.9 \leq m < 1$	76.41%	0.01636	512.79	70.47%	0.04577	1489.74	71.23%	0.13391	3891.11	74.2%	0.05083	1523
Call	$1 \leq m < 1.1$	73.04%	0.01766	568.72	72.3%	0.05104	1711.47	77.98%	0.1591	4760.26	74.23%	0.05983	1841.09
	$1.1 \leq m < 1.2$	82.95%	0.00841	270.36	74.2%	0.02593	917.51	76.8%	0.12482	3726.82	79.11%	0.05328	1602.69
	$1.2 \leq m < 1.3$	90.94%	0.005	153.64	77.8%	0.01441	508.12	77.69%	0.10568	3232.88	82.35%	0.05208	1606.86
	$1.3 \leq m < 1.4$	102.83%	0.00384	119.77	82.65%	0.00904	324.89	78.86%	0.0873	2713.84	85.74%	0.04951	1548.45
	$1.4 \leq m < 1.5$	113.55%	0.00313	98.23	88.11%	0.00631	224.26	80.37%	0.07434	2347.22	87.27%	0.04861	1539.96
	$1.5 \leq m < 1.6$	127.73%	0.00274	77.43	92.42%	0.00437	151.99	80.4%	0.05901	1863.78	87.82%	0.0418	1322.13
	$1.6 \leq m < 1.7$	140.46%	0.00254	71.72	99.01%	0.00394	131.02	82.55%	0.05407	1802.6	89.05%	0.04194	1397.31
	$1.7 \leq m < 2$	156.73%	0.00287	75.29	106.65%	0.00323	103.17	84.76%	0.04549	1470.9	89.38%	0.03946	1275.17
	$m \geq 2$	199.39%	0.00225	30.83	127.44%	0.00205	59.68	93.34%	0.02893	977.75	95.03%	0.02785	940.49
Total		89.09%	0.01115	351.82	87.69%	0.02097	704.31	84.6%	0.06797	2085.77	86.54%	0.04222	1308.34

CHAPTER 4

CRYPTO OPTION PRICING

Motivation: The bitcoin options market is highly fragmented and split between regulated and unregulated exchanges. This variety of different exchange types has created plain vanilla options with various settlement mechanisms, yet their implications are not widely understood. These different settlements influence not only the pricing of these financial instruments, but also the profit arising from crypto options trading. While stablecoin-denominated options might appear to be a safe alternative to fiat-denominated options, recent developments have indeed shown how unstable stablecoins can be. The risk associated with these necessitate the consideration of additional (exotic) options to bolster investor protection. Hence, a thorough comprehension of these products, as well as the ability to accurately price and hedge these is of utmost importance for all market participants.

Summary: We introduce and explain the settlement mechanisms of the three most traded bitcoin option types - standard, direct, and inverse bitcoin options - and derive sound mathematical expressions for their pricing. Additionally, we distinguish between crypto tokens categorised as securities or currencies and highlight the implications of such classifications for pricing methodologies. We show that inverse options can and should be priced the same way foreign exchange options are priced. Moreover, we underscore the risks associated with stablecoin-denominated options and introduce innovative derivatives designed to enhance risk management strategies and bolster overall market efficiency.

4.1 Underlying and Payoff

A token is a crypto asset that is transacted using a smart contract on a blockchain.[1] Almost all crypto assets are tokens, and even those that were minted before the introduction of Ethereum Request for Comments (ERC) token standards are easily 'wrapped' to become a token. Native tokens are generated through building the blocks in the chain. If the blockchain can carry smart contracts the native token is also the unit of account for the 'gas' required to fuel the smart-contract transactions.[2] Because a native token is used for payment, we can call it a cryptocurrency. Although the term 'cryptocurrency' has slipped into popular use as a generic term for all crypto assets, the only other type of token which is truly a cryptocurrency, i.e. a fungible token that is commonly used for payment and settlements, is a stablecoin. A stablecoin is a token whose price is pegged to a fiat currency such as the U.S. dollar.

Other tokens which are not used for payment can be divided into fungible and non-fungible classes. A non-fungible token (NFT) is a certificate of ownership of a physical or digital asset such as land or art. NFTs are frequently used as collateral for borrowing cryptocurrencies, but of course they themselves could not be a unit of account because they are non-fungible. An NFT is an asset which is arguably more akin to a commodity than a security. How to classify the non-payment but still fungible tokens that are issued by developers of projects in the digital economy is an even greater point of debate. Under the leadership of Gary Gensler and Dan Berkowitz, the SEC has been arguing for years that such tokens are a type of security, and should therefore be regulated by the SEC. On the other hand, the CFTC produced a counter-argument that they should be regulating the crypto assets, e.g. native tokens, that have a utility value and are therefore more like a commodity than a security.

Crypto assets are exchanged in trading pairs via decentralised liquidity pools on-chain or on centralised exchanges off-chain.[3] When both sides of the trading pair are cryptocurrencies, the trading pair is like an exchange rate. If neither side of the trading pair is a cryptocurrency, then the trading pair is more like an asset swap. And if only one side of the trading pair is a cryptocurrency, it is like a security with the cryptocurrency being the unit of account. The most common type of platform for trading exchange-rate crypto pairs is the order book of an off-chain centralised exchange such as Coinbase or Binance. By contrast, most asset-swap crypto pairs are traded in liquidity pools of an on-chain decentralised exchange, such as Uniswap or Curve. Most crypto-crypto asset swaps are traded in on-chain liquidity pools. Trading volumes in all these markets have exploded over recent years. In particular, millions

---

[1]This is the standard. However, other cryptos like IOTA or NANA do not use blockchains and operate independently.

[2]For example, ether, the native token of Ethereum is the unit of account for all non-fungible tokens minted onto Ethereum. And although Bitcoin is not a smart contract blockchain, its native token bitcoin is a common crypto unit of account for new token offerings.

[3]Or indeed an asset swap. Crypto-crypto asset swaps are most heavily traded in decentralised liquidity pools, recorded on blockchains, i.e. 'on-chain'. Not dissimilar to OTC agreements in traditional markets, except that on-chain transactions are fully transparent. But block production is rather slow so it could be a relatively long time before crypto-crypto on-chain swap derivatives are developed.

of retail traders, either truly believing in a projects' philosophy or for purely speculative purposes, have easily opened accounts on any number of self-regulated centralised exchanges to trade bitcoin and other crypto assets and derivatives. Large proprietary trading firms have also been trading very actively.[4] All these traders, on entering this melting pot of traditional finance and modern computer science, are faced with an array of innovative crypto assets many of which have very actively traded derivative products. However, in the crypto derivatives category, options are backed by very little academic research.

Apart from regulatory oversight, the main factor that differentiates the centralised crypto derivatives platforms is the type of products that they offer and their settlement. As discussed in the introduction, there are currently three distinct type of bitcoin options. For example, CME runs order books in standard European put and calls with a contract size in bitcoin or ether, and all contracts are margined and settled in USD. But ~90% of open interest and trading volume on crypto options has always been on Deribit, which only runs order books in so-called inverse options. These have a contract size of one bitcoin, ether or solana, and they track the USD value of these coins even though they are margined and settled in BTC, ETH or SOL. The cryptocurrency options on Binance complete the triplet of settlement mechanisms, as their options are order book direct options. That is, these options have an identical payoff to the standard options from the CME, but consider the bitcoin price in in a stablecoin like USDT or USDC. These three settlement mechanisms for crypto options account for all cryptocurrency option trades but are not yet widely understood. For instance, on Deribit the underlying is a non-tradable index of spot prices and on the CME it is a futures contract on a similarly non-tradable reference rate. The settlement price on Deribit is the average value of the underlying over the 30 minutes prior to settlement and on the CME it is the spot value of the reference rate. These settlement differences, combined with a widespread lack of proper documentation from the unregulated exchanges, may lead to confusion about a seemingly trivial European-style product.

This discussion has highlighted three important questions to answer before we can really understand crypto options should be priced and hedged:

1. What is the currency that the trader uses as a unit of account? This could be a cryptocurrency like BTC or a fiat currency such as USD. A BTC-based trader with a BTC-denominated trading account has a different perspective on profit and loss to a US trader whose account is measured in USD;

2. Is the underlying a security or a currency? If a security, then the option should be priced like a stock or bond option with the unit of account specified in 1, and if a currency then the option is equivalent to an FX option;

---

[4]Latest proprietary trading firms like Jane Street or Jump Trading are actively trading crypto, see Financial Times, accessed 25 September 2023.

3. Is the settlement price that of a tradable instrument? The vast majority of crypto options are settled using an index, sometimes also averaged prior to settlement. So should pricing and hedging take account of the incompleteness of this market or not?

Clarifying the answers to those questions justifies proposing that quanto options be added to the array of crypto products. Quanto direct options are similar to traditional quantos, but the quanto inverse option is a completely new type of exotic option. We argue that both direct and inverse quantos are better products for risk-averse USD-denominated agents than their vanilla counterparts which have no currency protection.

In this chapter, we discuss the properties of four different types of crypto options that are of practical and academic interest, with a specific focus on direct and inverse quanto products. We define the payoffs and settlements for direct and (quanto) inverse options precisely using USD as the fiat side and either BTC or ETH as the crypto side of the trading pair, while considering USDT as stabelcoin proxy for the direct and quanto options. Both bitcoin and ether are cryptocurrencies, but bitcoin is only a cryptocurrency while ether can also be regarded as a security or commodity because Ethereum is a smart contract blockchain whereas Bitcoin is not. Indeed, the SEC have argued that *every* token apart from bitcoin is a security and should therefore fall under the jurisdiction of the SEC rather than the CFTC.[5] We also consider two types of traders, one USD-denominated whose trading book is denominated in USD and the other crypto-based whose trading book is denominated in BTC.

4.2 Standard & Direct Options

Either standard or direct options are widely traded on all the regulated and some unregulated centralised exchanges. A bitcoin (ether) option trader can choose whether to trade the pair BTCUSD (ETHUSD) or BTCUSDT (ETHUSDT) or any other stablecoin such as USDC.[6] Because of the risks surrounding stablecoins (Duan and Urquhart, 2023), a USD-denominated trader may prefer the *standard* option, which is a plain vanilla European product. The call has payoff in USD given by:

$$V_T^\$ = \left(S_T^\$ - K^\$ \right)^+ , \tag{4.1}$$

where $K^\$$ denotes the strike price and $S_T^\$$ denotes the underlying price, both in USD, at maturity T. On the CME (at the time of writing) the underlying is a futures on either the

---

[5]In a recent SEC speech, Gary Gensler points out that "*[...] the vast majority [of crypto tokens] are securities*" but excludes bitcoin in particular in an earlier interview. The argument is that native token of other blockchains that are not smart-contract compliant, such as Dogecoin can be thought of a security because their primary use is as a 'meme' token. Just recently, the SEC announced to classify Ethereum and all its subsidiary tokens as a security. Thus, every project deployed on top of the Ethereum blockchain could be claimed to be a security and within SEC jurisdiction.

[6]The risks are huge. They include: market risk, because the stablecoin is only pegged to the value of USD and can deviate very far from the peg, in May 2022; but also operational risks (stablecoins are held on exchanges and so can be hacked); and regulatory risks (e.g. MiCA) directive places very firm caps on stablecoin trading volumes, to try to limit their capitalisation.

CME BRR – which from now on we regard as the BTCUSD rate as it tracks the FX rate, i.e. the number of USD for one unit of bitcoin – or the pair ETHUSD which we regard as the ETHUSD exchange rate or the USD price of the ETH security coin. The margining and settlement of standard bitcoin or ether options depends on the exchange. On the CME they are margined and settled in USD and the settlement price is that of the CME bitcoin or ether futures, so both the option premium and the payoff are denominated in USD. They are also cleared by the exchange, thus omitting counterparty risk. This procedure does not differ in any way from trading other commodities on the CME exchange. Binance lists *direct* BTCUSDT and ETHUSDT (and many other crypto) options which are margined and settled in USDT. The settlement price is that of the stablecoin-margined product BTCUSDT (or ETHUSDT) on their spot platform at the time of settlement, and the payoff is in USDT; but in other ways the Binance settlement procedure mimics the CME's. Hence, using the symbol $\mathbb{T}$ to denote the price of the stablecoin USDT, the call payoff is identical to (4.1) except the currency is USD not but the stablecoin USDT and may be written:

$$V_T^{\mathbb{T}} = \left(S_T^{\mathbb{T}} - K^{\mathbb{T}}\right)^+. \tag{4.2}$$

We have selected these two particular centralised exchanges because they are the largest to list standard and direct crypto options, and they illustrate how these exchanges can offer the same basic product, but have different underlyings and settlement procedures. The primary distinction between *standard* and *direct* options depends on whether they are fiat- or stablecoin-denominated. The main motivation to distinguish between these two types arises from the instability of stablecoin as Duan and Urquhart (2023) highlight and which we shall discuss in-depth in the following section.

Determining the pricing approach for standard or direct options is fundamentally linked to the trader's interpretation of the underlying asset, i.e. bitcoin or ether. In circumstances where the trader classifies these as currency pairs, the payoff structure of these assets allows a direct pricing approach using the Garman and Kohlhagen (1983), or any other FX pricing model. However, the question on the risk-free rate arises as there is no governmental body issuing treasury notes for any cryptocurrency. Further, lending platforms that guarantee fixed annual returns bear substantial risk, as we shall discuss in the next section. On the other side of the spectrum, following the SEC's announcement, treating any crypto token as security enables the use of standard option pricing models like the Black and Scholes (1973), or other SV model. The challenge here lies in identifying elements comparable to dividend payouts. While the proof-of-work architecture of bitcoin does not pay any reward, the proof-of-stake structure of Ethereum allows investors to deposit, or stake, their ETH and collect a weekly return. Alternatively, investors possessing a sizable amount of ETH might acquire a form of dividend through validation of transactions, leading to a pricing scenario dependent on the wealth of the individual agent. Irrespective of interpretation due to the conventional, straightforward payoffs, there exists a plethora of available option pricing models.

Figure 4.1 illustrates the well-known payoffs of standard and direct call and puts as a function of the underlying. For the sake of clarity, we omit the graphs for USDT as these would not add any further information. It shows the well-know piece-wise linear payoff of a plain vanilla call otpion.

Figure 4.1: Standard and Direct Option Payoff

The payoff to a long call (left; blue graph) and put (right; red graph) as a function of the settlement price. These payoffs present a standard option but could also display the direct payoff using USDT ($\mathbb{T}$) instead of USD ($\$$) for both the payoff (vertical axis) as well as the underlying (vertical axis). The strike level is set at $K^{\$} = \$25,000$.

4.3 Inverse Options

Inverse options are the only type of option Deribit lists. Therefore, over 85% of the trading volume on centralised crypto options exchanges is on inverse options. The reason Deribit only lists inverse options is that it is a non-fiat exchange so there is no USD transacted anywhere on the platform, although Deribit indicated to enable fiat onbording in the future without naming an exact date. The inverse structure allows Deribit to list options on cryptocurrency-USD trading pairs, and each option is margined and settled in the option's cryptocurrency.

Many other non-fiat exchanges list inverse options, precisely because they can trade against the USD without using it as the unit of account or indeed allowing any fiat currency onto the platform. And most inverse options track the USD value of a coin index not a single spot price. Importantly, margining and settlement is always in a cryptocurrency, not in fiat. If held to maturity, the settlement price S_T is either the coin index value exactly at the settlement time, or its average value over a time period immediately prior to settlement. Due to its frequent rebalancing of the Deribit bitcoin index, physical replication of this index is an immensely difficult and expensive task. An agent would be required to hold bitcoin positions on multiple exchanges and rebalance these constantly. Complicating replication, even more,

the final option settlement value is the average of the index during the last 30 minutes before expiry. This important feature about Deribit inverse options is often left out. The underlying is not directly tradable. For this, we must consider pricing within an incomplete market.

4.3.1 The Inverse Structure

In general, an inverse contract specifies a notional number N of coins which is multiplied by a point value to obtain a payoff expressed as a number of coins, i.e. in the units of the coin. The terminal payoff (and indeed all trading profits) are transferred to the trader in the cryptocurrency, not in USD. For example, the payoff $V_T^{\mathbb{B}}$ to an inverse call on BTCUSD is denominated in BTC and may be written:

$$V_T^{\mathbb{B}} = N \, \frac{\left(S_T^{\$} - K^{\$}\right)^+}{S_T^{\$}}, \tag{4.3}$$

where the second term is a dimensionless quantity which we denote the point value. There is actually no need for our notation to specify the units for the settlement price and strike (even though we have done so above) because their difference becomes dimensionless when divided by $S_T^{\$}$. But we do need to know what the underlying is, e.g. the BTCUSD rate.

Currently, all exchanges that list inverse options use a notional of exactly one coin, i.e. Deribit set $N = 1$ bitcoin, or 1 ether or 1 sol, depending on the underlying. Moreover, they all quote the option prices in USD, as well as in the cryptocurrency of the trading pair, but settlement is always in the cryptocurrency. Now, for a USD-denominated trader the true payoff (which is in cryptocurrency) may just as well be translated into USD, in which case $V_T^{\mathbb{B}}$ should be multiplied by the price of the underling at the time of settlement $\bar{S}_T^{\$}$. Note that this price is not the same as the settlement price, the latter being the average price over the 30 minutes before the settlement time. Given how volatile crypto markets are, there can be a large difference. Anyway, we can write the payoff to a USD-denominated trader as:

$$V_T^{\$} = \bar{S}_T^{\$} V_T^{\mathbb{B}} = \bar{S}_T^{\$} \frac{\left(S_T^{\$} - K^{\$}\right)^+}{S_T^{\$}} \approx \left(S_T^{\$} - K^{\$}\right)^+.$$

This shows that one can think of the inverse option payoff as equivalent to a standard FX option, except that the payoff is denominated in the foreign currency. There is a large body of academic research on FX options, their pricing, hedging, volatility dynamics and so forth, see Levy (1992), Carr and Wu (2007), and Demeterfi (1998) and many others. However, FX options are usually denominated in the same currency as the underlying. Inverse options are denominated in the foreign rather than the domestic currency and this rather unusual denomination of the payoff is a potential source of confusion. Similar characteristics in FX options, when one uses the foreign-domestic symmetry relationship to convert a domestic call to a foreign put, has been documented previously (Grabbe, 1983; Reiswich and Wystup, 2010). We also provide more details on the foreign-domestic relationship in Section 4.3.2.

Figure 4.2 illustrates the true terminal payoff an agent would receive when trading a long inverse option. It is a piece-wise concave (call) or entirely convex (put) structure. Note that the call payoff is capped at maximum ₿1 whereas the put can theoretically pay out an infinite amount of bitcoin. For USD-denominated traders, but only for these traders, the payoff to an inverse option can be approximated by a standard piece-wise linear form,[7] and hence their pricing is almost identical to that of the direct option (Garman and Kohlhagen, 1983), as seen in the next subsection. But very much depends only on the trader's base currency. USD-denominated traders might only consider the USD value of their option position on the balance sheet, but crypto option traders are international, and so may prefer to use a cryptocurrency – or a different fiat currency – as their unit of account on Deribit. And even for USD-denominated traders, cryptocurrency-denominated profits stemming from large positions in inverse options cannot be instantly exchanged for USD, which adds a liquidity premium the inverse Deribit options. This is a particular issue for puts which can, in theory, pay out an infinite amount of bitcoin. Deribit also offers block trades of at least 25 bitcoin or 250 ether for which it does not require very large negative price movements for the position holder to receive a high number of bitcoin or ether. Thus, while the *paper* value of profits in BTC or ETH may be very high, converting such an amount into fiat may well result in liquidity problems. Then profits would be reduced through spillage on the currency conversion trades. The absence of adequate brokerage services, and of exchange requirements to hold large margin reserves, are yet another expense factor for institutional bitcoin option trading. Professional traders transacting large amounts might refrain from denominating their profit and loss in USD and use cryptocurrency their balance sheets instead, but still the currency risk faced traders in inverse options cannot be ignored.

4.3.2 Complete Market Pricing

This subsection examines the true pricing of inverse options under the assumption that the underlying is tradable and the market complete. For this case, we consider bitcoin as a currency, and the underlying of a bitcoin option contract is the BTCUSD exchange rate.[8] We denote the price of the tradable underlying, the BTCUSD exchange rate, at time t as $S_t^\$$. Assuming well-functioning money markets exist for both currencies, we can define two cash bond accounts $B_t^\$ = e^{r^\$ t}$ and $B_t^\mathbb{B} = e^{r^\mathbb{B} t}$ as the respective numéraires, where $r^\$$ and $r^\mathbb{B}$ are the risk-free interest rated in the corresponding currencies. Furthermore, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a filtered probability space with filtration $\mathcal{F}$ and probability measure $\mathbb{P}$. Following Clark

---

[7]Approximate because $\bar{S}_T \neq S_T$, unless there is no trading in the 30 minutes before expiry. However, this difference affects only the payoff and not the fair price at any time prior to expiry.

[8]Note that this is just an experimental framework. There is no major exchange offering these exact options except for maybe Binance. However, the Binance direct option is on the BTCUSDT exchange. Given the recent events involving the stable coin Terra, we want to point out possible risk factors involving the use of supposedly stable stable coins without discussing their benefit of existing and refer the reader to Baur and Hoang (2021) or Hoang and Baur (2021). Any other interpretation of the underlying, i.e. as a security, is also possible as this does not change the theoretical derivation or the change of numéraires.

Figure 4.2: Inverse Option Payoff

The payoff in bitcoin to a long inverse call (left; blue graph) and put (right; red graph) as a function of the settlement price. The strike is $25,000. Note that the call payoff is capped at maximum ₿1 whereas the put can theoretically pay out an infinite amount of bitcoin.

(2011), we denote the dynamics of the log-normal asset $S_t^{\$}$ under the physical measure $\mathbb{P}$ as:

$$\frac{dS_t^{\$}}{S_t^{\$}} = \mu^{\$} dt + \sigma^{\$} dW_t^{\$},$$

where $\mu^{\$}$ is the drift, $\sigma^{\$}$ is the volatility, and $W^{\$}$ denotes a standard Brownian motion. In this framework, we can rewrite the payoff $V_T^{\mathbb{B}}$ at maturity T of a call on $S_T^{\$}$ from (4.3) as:

$$V_T^{\mathbb{B}} = K^{\$} \left(K^{\mathbb{B}} - S_T^{\mathbb{B}} \right)^+,$$

where $K^{\mathbb{B}} = \frac{1}{K^{\$}}$ and $S^{\mathbb{B}} = \frac{1}{S^{\$}}$ is the opposite exchange rate, i.e. USDBTC. Apply Itô's lemma, $S_t^{\mathbb{B}}$ is governed by:

$$\frac{dS_t^{\mathbb{B}}}{S_t^{\mathbb{B}}} = \mu^{\mathbb{B}} dt - \sigma^{\mathbb{B}} dW_t^{\$},$$

under $\mathbb{P}$, where $\mu^{\mathbb{B}} = (\sigma^{\$})^2 - \mu^{\$}$ and $\sigma^{\mathbb{B}} = \sigma^{\$}$. We further define a Equivalent Martingale Measure (EMM) $\mathbb{Q}$ with the Radon-Nikodym density:

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = M_T, \quad \text{with} \quad M_t = \exp\left[-\int_0^t m_u dW_u^{\$} - \frac{1}{2} \int_0^t m_u^2 du \right],$$

where $\{M_t\}_{0 \leq t \leq T}$ is a $\mathbb{P}$-martingale. Applying the Skorokhod reflection principle and the Girsanov theorem, the process:

$$\tilde{W}_t^{\mathbb{B}} = -W_t^{\$} + \int_0^t m_u du$$

is a $\mathbb{Q}$-Brownian motion, and $S_t^{\mathbb{B}}$ is governed by:

$$\frac{dS_t^{\mathbb{B}}}{S_t^{\mathbb{B}}} = \left(\mu^{\mathbb{B}} - m_t \sigma^{\mathbb{B}}\right) dt + \sigma^{\mathbb{B}} d\widetilde{W}_t^{\mathbb{B}},$$

under $\mathbb{Q}$. Requiring $S_t^{\mathbb{B}} B_t^{\$} / B_t^{\mathbb{B}}$ to be a $\mathbb{Q}$-martingale, we can rewrite:

$$m_t = \frac{\mu^{\mathbb{B}} - \left(r^{\mathbb{B}} - r^{\$}\right)}{\sigma^{\mathbb{B}}}.$$

It follows that under $\mathbb{Q}$ the dynamic of $S_t^{\mathbb{B}}$ is given by:

$$\frac{dS_t^{\mathbb{B}}}{S_t^{\mathbb{B}}} = \left(r^{\mathbb{B}} - r^{\$}\right) dt + \sigma^{\mathbb{B}} d\widetilde{W}_t^{\mathbb{B}}.$$

Note that a USD-denominated call becomes a BTC-based put with strike $K^{\mathbb{B}}$ being the inverse of the USD-strike price. The option can be priced trivially via standard FX option pricing formulae, and satisfies the put-call duality conditions as articulated, for example, in Garman and Kohlhagen (1983). The price of the put in BTC is given by:

$$P_t^{\mathbb{B}} = e^{-r^{\mathbb{B}}\tau} K^{\mathbb{B}} \Phi\left(-d_2^{\mathbb{B}}\right) - e^{-r^{\$}\tau} S_t^{\mathbb{B}} \Phi\left(-d_1^{\mathbb{B}}\right) \tag{4.4}$$

$$d_1^{\mathbb{B}} = \frac{1}{\sigma^{\mathbb{B}}\sqrt{\tau}} \left[\ln\left(\frac{S_t^{\mathbb{B}}}{K^{\mathbb{B}}}\right) + \left(r^{\mathbb{B}} - r^{\$} + \frac{(\sigma^{\mathbb{B}})^2}{2}\right)\tau\right]$$

$$d_2^{\mathbb{B}} = d_1^{\mathbb{B}} - \sigma^{\mathbb{B}}\sqrt{\tau}.$$

Similarly, we can express the variables in an USD-denominated framework:

$$d_1^{\mathbb{B}} = \frac{1}{\sigma^{\$}\sqrt{\tau}} \left[-\ln\left(\frac{S_t^{\$}}{K^{\$}}\right) - \left(r^{\$} - r^{\mathbb{B}} - \frac{(\sigma^{\$})^2}{2}\right)\tau\right] = -d_2^{\$}$$

$$d_2^{\mathbb{B}} = -d_1^{\$}.$$

It becomes obvious that $K^{\$}$ units of BTC-denominated puts are equivalent to $1/S_t^{\$}$ units of USD-denominated calls:

$$K^{\$} \ P_t^{\mathbb{B}} = e^{-r^{\mathbb{B}}\tau} \Phi\left(-d_2^{\mathbb{B}}\right) - e^{-r^{\$}\tau} K^{\$} S_t^{\mathbb{B}} \Phi\left(-d_1^{\mathbb{B}}\right) = \frac{1}{S^{\$}} C_t^{\$}. \tag{4.5}$$

This pricing approach can be readily extended beyond Black-Scholes world. Solutions to option pricing problems are attainable, at least in the Fourier transform sense, for any tractable Lévy processes governing the evolution of the USDBTC exchange rate. Popular choices are the Heston model, the SABR model and their multi-factor and/or jump-extended variants, among others (Sepp and Rakhmonov, 2022).

4.3.3 Incomplete Market Pricing

The simple pricing approach in Section 4.3.2 relies on the assumptions that both markets denominated in USD and in BTC are complete, and there are no restrictions on exchanging wealth from one to the other. To be most precise, and for the sake of fully understanding the Deribit options market, we must note that this assumption is not satisfied. Not only is there no well-functioning money market for bitcoin, it does not exist at all. In fact, lending platforms are the closest that come to some kind of money market. These companies promise risk-free interest for clients willing to deposit their crypto. However, poor risk management and a dangerous mixture of incompetence and ignorance brought major lenders to their knees just recently.[9] Second, the exchanges on which inverse options are traded (which accept no fiat and are unregulated) use non-traded underlyings. For example, the Deribit bitcoin inverse options use their own index as the underlying, which is an average of bitcoin spot prices from (currently) four, but can reach up to eleven different centralised exchanges. Due to its frequent rebalancing and arbitrary exclusion of exchanges, the physical replication of this index is an immensely difficult and expensive task. An agent would be required to hold bitcoin positions on multiple exchanges and rebalance these constantly. Moreover, the final option settlement price is the average of the index during the last 30 minutes before expiry. This important feature about Deribit inverse options is often ignored. But the price of bitcoin can change considerably during 30 minutes – much more than we see for traditional financial instruments. Hence, the underlying is not tradable and the market is incomplete. Therefore, equation (4.4) provides only an approximate option price.

To fully understand the valuation and hedging of bitcoin options on Deribit, we need to consider the actual state of the cryptocurrency option market. To circumvent the difficulties in reconstructing and trading the underlying index of the Deribit bitcoin options we consider instead the Deribit perpetual futures, because this tracks the index closely through the funding mechanism. However, this introduces an unhedgeable basis risk that can be sizable in a volatile market, as seen in Figure 3.2. Now, option pricing in an incomplete market in the presence of basis risk is typically solved through indifference pricing by formulating a stochastic control problem in the mean of the Hamilton-Jacobi-Bellman (HJB) partial differential equation (Monoyios, 2004), or by solving the corresponding forward-backward stochastic differential equation (Rouge and El Karoui, 2000). Consider the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ as in the standard inverse pricing case. The non-tradable underlying asset (in this case the DBTC, $S_t^\$$)

---

[9]Celsius and Voyager are the latest big players filing for bankruptcy shattering trader's trust in the whole crypto ecosystem.

and the hedging instrument (the perpetuals, $F_t^{\$}$) evolve according to the GBMs:

$$\frac{dS_t^{\$}}{S_t^{\$}} = \mu^{\$}dt + \sigma^{\$}dW_t^{\$}, \tag{4.6}$$

$$\frac{dF_t^{\$}}{F_t^{\$}} = \mu dt + \bar{\sigma}dW_t, \tag{4.7}$$

where $W^{\$}$ and W correlated standard Brownian motions with $\langle dW^{\$}, dW \rangle = \rho \, dt$. Note that the second fundamental theorem of mathematical finance describes the ability to hedge arbitrary claims; or, put another way, it states that an arbitrage-free market is complete if and only if the martingale measure corresponding to the numéraire is unique. Thus, as S and F are not perfectly correlated, i.e. $\mid \rho \mid < 1$, we cannot find a unique EMM for which the discounted value of F is a martingale under risk-neutral measure and consider the market incomplete.[10] We want to highlight that the correlation between the non-tradeable index and tradeable perpetual is very high (> 0.99) over the entire time and this discussion is rather theoretical in nature and serves to fully understand the market. We rewrite

$$W_t^{\$} = \rho W_t + \epsilon \tilde{W}_t^{\$},$$

where $\epsilon = \sqrt{1-\rho^2}$ with independent $\tilde{W}_t^{\$}$ and W_t. We further denote by $\{\mathcal{G}_t\}_{0 \le t \le T}$ the filtration generated by $W_t^{\$}$. Note that this Brownian motion drives the non-traded inverse index asset. Without losing generality, we assume the dividend yield is zero, and the US money market provides a risk-free interest rate r. Assume there exists an equivalent measure $\mathbb{Q}$ to $\mathbb{P}$ on $\mathcal{F}$. Then there exists adapted processes m_T and g_T where the Radon-Nykodym derivative is given by:

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = M_T,$$

where $\{M_t\}_{0 \le t \le T}$ is a $\mathbb{P}$-martingale with M_t is given by:

$$M_t = \exp\left[\int_0^t m_u dW_u + \int_0^t g_u d\tilde{W}_u^{\$} - \frac{1}{2}\int_0^t m_u^2 du - \frac{1}{2}\int_0^t g_u^2 du \right].$$

Using the multidimensional Girsanov theorem we see that the processes $\left\{ \tilde{W}_t, \hat{W}_t^{\$} \right\}_{0 \le t \le T}$ defined by:

$$\begin{pmatrix} \tilde{W}_t \\ \hat{W}_t^{\$} \end{pmatrix} = \begin{pmatrix} W_t + \int_0^t m_u du \\ \tilde{W}_t^{\$} + \int_0^t g_u du \end{pmatrix},$$

---

[10]Note that a market can be incomplete in many ways. Some reasons include: (i) more random sources than risky assets present, (ii) constraints on admissible positions, (iii) underlying object is not traded, (iv) iliquid market or (v) positions cannot be carried forward in time without large costs. For a detailed mathematical definition see for example Björk (2020).

is an independent Brownian motion under $\mathcal{Q}$. Further, the dynamics under $\mathcal{Q}$ are given by:

$$\frac{dS_t^\$}{S_t^\$} = \left(\mu^\$ - \sigma^\$ \left(\rho m_t + \epsilon g_t\right)\right) dt + \sigma_\$ dW_t,$$

$$\frac{dF_t^\$}{F_t^\$} = \left(\mu - \bar{\sigma} m_t\right) dt + \bar{\sigma} d\hat{W}_t,$$

where $\hat{W}_t = \rho \widetilde{W}_t + \epsilon \hat{W}_t^{\mathbb{B}}$ is a Brownian motion such that $\langle d\widetilde{W}_t, d\hat{W}_t \rangle = \rho\, dt$. For $\mathbb{Q}$ be be a local EMM, $F_t^\$$ needs to be a $\mathbb{Q}$-local martingale, i.e. iff:

$$\mu - m_t \bar{\sigma} = r \quad \Rightarrow \quad m_t = m := \frac{\mu - r}{\bar{\sigma}}.$$

Note that we assume that the model is generically arbitrage free, that is m_t, viewed as a linear mapping function, is surjective $\mathbb{P}$-a.s. and that the EMM is uniquely defined. On the other hand $S^\$$ is non-tradable which lets g_T be of any arbitrary form which results in an infinite set of possible EMM. We define this set as $\mathcal{M}$ which is in correspondence with the set of g_t. We further want to link the local martingale with an equivalent probability measure. Denote an equivalent measure $\widetilde{\mathbb{P}}$ to $\mathbb{P}$ on $\mathcal{G}$, the risk-neutral density given by:

$$\frac{d\widetilde{\mathbb{P}}}{d\mathbb{P}} = \widetilde{M}_T, \quad \text{with} \quad \widetilde{M}_t = \exp\left\{ -\int_0^t \theta_u dW_u^\$ - \frac{1}{2} \int_0^t \theta_u^2 du \right\},$$

where θ_t is a $\mathcal{G}_t$-adapted process. Under $\widetilde{\mathbb{P}}$ we define $\hat{W}$ as:

$$\hat{W}_t = W_t^\$ + \int_0^t \theta_u du$$

and underlying dynamics under $\widetilde{\mathbb{P}}$ given by:

$$\frac{dS_t^\$}{S_t^\$} = \left(\mu^\$ - \sigma^\$ \theta_t\right) dt + \sigma^\$ d\hat{W}_t .$$

Note that the dynamics of the non-tradable index are the same under $\mathbb{Q}$ and $\widetilde{\mathbb{P}}$ when the integrands m_t, g_t and θ_t are given as:

$$\rho m_t + \epsilon g_t = \theta_t.$$

Many arbitrage valuation models such as the BS model rely on the assumption that any option's claim can be replicated by a portfolio and hence hedged perfectly. The absence of such a hedging portfolio leads to major difficulties when evaluating options in this fashion. Any further derivation would require alternative methods to handle the pricing in incomplete markets. These could include (i) a restriction to smaller Girsanov transformation such as the generalised Escher transformation (Kallsen and Shiryayev, 2002), (ii) the introduction of

a generalised distance between probability measures and choose the EMM which minimises the distance such as the concept of f-divergence (Goll and Rüschendorf, 2001), (iii) through the duality portfolio optimisation (Schachermayer, 2002), rather than finding a unique martingale measure, find a 'reasonable' measure (Björk and Slinko, 2006), or (v) using a specific preference-dependent approach of the agent. Particularly the latter one enjoys great popularity in financial literature. By making assumptions about the shape of the trader's utility function, the option pricing problem could be transformed into an optimal trading strategy problem where the trader seeks to maximise his expected utility with and without trading in options contracts. Many papers already address this topic in much detail – see Rouge and El Karoui (2000), Monoyios (2004), Ankirchner et al. (2010) and Davis (2000) and many others. This way it could be presented as the numerical solution to the corresponding HJB partial, or the forward-backward stochastic, differential equation. However, there are significant limitations to using this type of model for pricing and hedging purposes. The option pricing formula and Greeks depend on whether the option position is long or short, and the type of utility functions used can often be restrictive, in some cases it would even yield option prices that depend on the agent's initial wealth.

However, there exists a trade-off between the complete and the incomplete market model, i.e. facing potential pricing error due to basis risk versus the necessity of alternative pricing methods. On the one hand, Figure 3.2 shows that the basis risk between index and perpetuals oscillates mostly around ± 10 bps, particularly since 2022. On the other hand, specifically for the preference dependent pricing, it could be challenging to identify the correct preferences due to prominent sentimental and behavioural features in crypto markets, specifically, due to the diversity of traders, i.e. (semi-professional) retail and institutional participants. Formulating correct assumptions of the agent's preference is likely unrealistic, as this space is continuously evolving, and the discussion in Section 3.2 – specifically, the open interest for deep OTM options ($m > 2$) seen in Table 3.1 – leads us to question the rationality of some bitcoin option traders. Thus, the risk of mis-specifying investor preference outweighs the basis risk. Further analysis may be warranted to discern the sentimental and behavioural features, but will not be addressed in this thesis.

4.4 Quanto Direct Options

A quantity-adjusted option, or quanto option for short, allows traders to gain exposure to a foreign market without taking any currency risk. In traditional markets the underlying is often a single security, or a security index, or another asset like a commodity. The settlement price S_T of this asset and the option strike K are denominated in the foreign currency but the standard option payoff is converted into domestic currency using a predetermined exchange rate $\bar{X}$ which is agreed upon entering the (usually OTC) contract. For instance, a standard

quanto call has payoff:

$$V_T = \bar{X} \, (S_T - K)^+ . \tag{4.8}$$

Because they are well-known products in traditional markets the pricing and hedging for quanto options has been researched very extensively (Jamshidian, 1993; Demeterfi, 1998; Jeanblanc et al., 2009; Clark, 2011). So it is somewhat surprising that there has been no previous research documenting quanto products in crypto markets. In fact, the first futures ever traded on crypto were quanto products.[11]

The main cryptocurrencies like bitcoin and ether have very active options markets but there are thousands of minor tokens that are only paired with stablecoins on the major non-fiat exchanges like Binance, never with USD. Therefore a USD-denominated trader wishing to trade such a token – let us call it XYZ – must take on the (very real) risk of the stablecoin de-pegging from USD while the trade on XYZ is in place, as seen with the algorithmic stablecoin TerraLuna. (Duan and Urquhart, 2023) are among the first to examine the stability of the five largest, currently available stablecoins and indeed find strong evidence of instability. To remove this currency risk, at the same time as leveraging their exposure to XYZ through an option trade, a USD-denominated agent might agree a fixed stablecoin exchange rate with the quanto option issuer. For instance, using a fix of the tether rate $\bar{X}^{\mathbb{T}/\$}$ the quanto direct call USD-denominated payoff becomes:

$$V_T^\$ = \bar{X}^{\mathbb{T}/\$} \left(S_T^\mathbb{T} - K^\mathbb{T} \right)^+ , \tag{4.9}$$

where $S_T^\mathbb{T}$ is the tether price of the XYZ token at the time of the option maturity T and the option strike $K^\mathbb{T}$ of the quanto is also denominated in USDT, although any other stablecoin such as USDC could replace tether in this payoff. To illustrate the usefulness of a quanto option to crypto traders further, first suppose a BTC-based option trader is interested to gain exposure to ETH while keeping his BTC position. He could use existing exchange-traded products to convert BTC to USD and then trade ETHUSD options, facing transaction costs and crossing the spread twice. Alternatively, the trader can obtain a payoff denominated in BTC by agreeing a fixed USDBTC rate $\bar{X}^{\$/\mathbb{B}}$ with the quanto option issuer before entering the contract on ETHUSD. We express the payoff to a quanto call, for a BTC-based trader as:

$$V_T^\mathbb{B} = \bar{X}^{\$/\mathbb{B}} \left(S_T^\$ - K^\$ \right)^+ , \tag{4.10}$$

where $S_T^\$$ is the price of ETH in USD at maturity T and the option strike $K^\$$ is also denominated in USD. These types of options would enable traders to participate in ETH without physically owning or depositing ether, thus avoiding gas or other fees. Figure 4.3 depicts the payoffs to a quanto call and put given by (4.10).

---

[11]The derivatives exchange BitMEX was among the first exchanges to offer quanto products in form of futures on various coin/coin or coin/USDT pairs, see BitMEX quanto futures.

Figure 4.3: Standard Quanto Option Payoffs

The payoffs to a long standard quanto call (left; blue graph) and put (right, red graph). The security (in this case) is ether, the foreign currency is USD and the domestic currency is BTC, i.e. the buyer is a BTC-based trader. Thus, the payoff is in BTC and the underlying on the horizontal axis is the USD price of ETH. We set $K^{\mathbb{E}} = \$1750$.
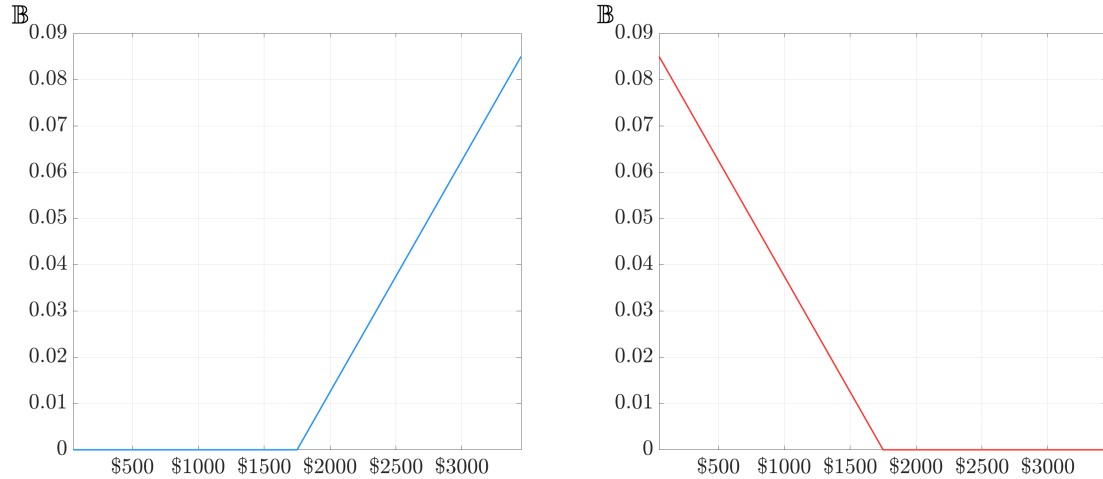
4.5 Quanto Inverse Options

A quanto inverse option is a natural extension of both inverse and quanto direct options which converts the inverse option payoff to another currency using an exchange rate that is fixed upfront and agreed between both parties when the contract is issued. This allows traders to mitigate the currency risk that is unavoidable when trading inverse options.

4.5.1 Structure

The payoff to a quanto inverse call with a notional of $N = 1$ coin in cryptocurrency $\mathbb{Y}$ can be converted to a currency $\mathbb{Z}$ (either crypto or fiat) at a fixed exchange rate $\bar{X}^{\mathbb{Y}/\mathbb{Z}}$, yielding the payoff:

$$V_T^{\mathbb{Z}} = \bar{X}^{\mathbb{Y}/\mathbb{Z}} \, \frac{\left(S_T^{\mathbb{Z}} - K^{\mathbb{Z}}\right)^+}{S_T^{\mathbb{Z}}}, \tag{4.11}$$

where $S_T^{\mathbb{Z}}$ and $K^{\mathbb{Z}}$ are the settlement and strike prices of an option on a token XYZ, both denominated the cryptocurrency $\mathbb{Z}$. These options have payoffs that mimic the convex put and concave call payoffs of inverse options, but they are denominated in a different currency.[12] The quanto factor $\bar{X}^{\mathbb{Y}/\mathbb{Z}}$ changes the slope of the terminal payoff and consequently also the option price prior to expiry. Depending on the choice of $\bar{X}^{\mathbb{Y}/\mathbb{Z}}$, a quanto inverse call (or put) could have higher or lower prices than a direct call (or put), as shown later. Intuitively, the buyer of such a call may seek to fix the exchange rate slightly higher or lower to his expected

---

[12]In fact, the quanto inverse option has a similar intuition to the quanto clique option (Mercurio, 2003) or the amortising option (Gatheral, 2006).

future exchange rate, depending on his position. Compared with their direct counterpart, a quanto inverse call loses more value as the option moves further in the money. As such, a quanto inverse call provides an affordable alternative to gain exposure to a crypto option by paying a lower price than one would for an ITM direct call of the same strike.

Figure 4.4 displays the payoffs to quanto inverse calls and puts on the BTCUSD trading pair. In this example we also use a fixed BTCUSD exchange rate (but in general the exchange-rate fix could be in BTC relative to any other crypto or fiat currency) so the shapes of these payoffs are exactly similar to the non-quanto inverse option payoff in Figure 4.2. The only difference is that the vertical axis is now in USD units. In particular, the quanto inverse put payoff again increases rapidly as the underlying depreciates. This feature makes quanto inverse puts an excellent insurance against a black swan crypto event. For instance, suppose that in February 2020 a trader bought a quanto inverse put, with a notional of $\mathbb{B}1$, expiring on 13 March 2020. Also suppose both the strike and the quanto factor of this put were fixed at $9000, which is reasonable since this was about the BTCUSD rate in February 2020. On 12 March 2020 the BTCUSD rate fell to about $3500, so suppose this was the settlement price. Thus a standard put would have paid $5500 if held to expiry. But the quanto inverse put would have paid $14,143, which is more than double that of the standard put. Finally, consider how a quanto inverse option might be constructed for a USD-denominated trader who seeks exposure to a crypto asset XYZ, but instead of trading a direct option on XYZ with currency protection against decoupling of stablecoin price from USD, as in Section 4.2, the trader prefers to gain exposure to an inverse option payoff. Reasons for this could be: (i) for the same terminal value of the underling, the quanto inverse option profit can be greater than that from direct option; and (ii), a quanto inverse put option payoff is convex and uncapped.

Figure 4.4: Quanto Inverse Options

The payoff to a long quanto inverse call (left; blue graph) and put (right; red graph) as a function of the settlement price. In (4.11) we set $\mathbb{Y} = \mathbb{B}$ i.e. the notional is 1 bitcoin, and translate the inverse option payoff to USD by setting $\mathbb{Z} = \$$. Both the quanto factor and the option strike are set to $25,000.

4.5.2 Pricing

It's worth noting upfront that the quanto inverse option shares a payoff structure similar to the inverse option payoff, but is denominated differently. This subtle difference has important consequences in terms of how these options should be priced. Consider first a standard FX call on BTCUSD that matures at time T, with either a USD-denominated payoff:

$$V_T^{\$} = \left(S_T^{\$} - K^{\$}\right)^+,$$

or a BTC-denominated payoff:

$$V_T^{\mathbb{B}} = K^{\$}\left(K^{\mathbb{B}} - S_T^{\mathbb{B}}\right)^+.$$

We have already shown that this option can be priced in a standard FX option pricing framework. Now consider an exotic option that pays:

$$V_T^{\$} = \bar{X}^{\mathbb{B}/\$}\frac{\left(S_T^{\$} - K^{\$}\right)^+}{S_T^{\$}}, \tag{4.12}$$

where $\bar{X}^{\mathbb{B}/\$}$ is a predetermined exchange rate that transforms the bitcoin denominated point value of an standard inverse option to a USD payoff. In the following we simplify notation and omit the superscript, so we set $\bar{X}^{\mathbb{B}/\$} = \bar{X}$. Similar to CME options, these would be USD-settled options, on the exchange rate of any token, or a reference rate or a futures contract. In this example, we focus on the USD value of one bitcoin. Using the same notation as in Section 4.3.2 we denote the underlying of this product at maturity by S_T. Note that the underlying, as well as the strike are both denominated in USD. However, for the sake of clarity and readability, we shall omit any superscript for these variables, as well as for the option's payoff and risk-free rate, as they all follow a USD denomination in this subsection. To avoid any confusion between the volatility driving the GBM at hand, and shall denote $\sigma = \sigma^{\$}$. Without losing generality, we assume the dividend yield is zero, and the US money market provides a risk-free interest rate r. The risk-neutral $\mathbb{Q}$-dynamics of the underlying are given by:

$$\frac{dS_t}{S_t} = rdt + \sigma dW_t,$$
$$S_T = S_t \exp\left\{\left(r - \frac{1}{2}\sigma^2\right)(T - t) + \sigma(W_T - W_t)\right\}.$$

Now let $Y_t = (S_t)^{-1}$. By Itô's lemma the $\mathbb{Q}$-dynamics of Y_t are:

$$\frac{dY_t}{Y_t} = \left(\sigma^2 - r\right)dt - \sigma dW_t.$$

The value of the discounted price process is therefore:

$$\tilde{Y}_t = e^{-r\tau}\, \mathbb{E}^{\mathbb{Q}}\left[Y_T^{\$} \mid \mathcal{F}_T\right] = Y_T e^{(\sigma^2 - 2r)\tau},$$

where $\tau = T - t$. We note that $\tilde{Y}_t$ is a martingale under the $\mathbb{Q}$-measure and hence we perform a change of numéraire from the risk-neutral measure $\mathbb{Q}$ to an equivalent martingale measure $\tilde{\mathbb{Q}}$ where $\tilde{Y}_t$ is the new numéraire. We can rewrite the dynamics of $S_t^{\$}$ as:

$$\frac{dS_t}{S_t} = \left(r - \sigma^2\right) dt + \sigma d\tilde{W}_t,$$

$$S_t = S_0 \exp\left\{\left(r - \frac{3}{2}\sigma^2\right) t + \sigma \tilde{W}_t\right\},$$

where $\tilde{W}_T = W_t + \sigma t$ is a Wiener process under the martingale measure $\tilde{\mathbb{Q}}$. Under these assumptions, we can express the Radon-Nikodym derivative by:

$$\left.\frac{d\tilde{\mathbb{Q}}}{d\mathbb{Q}}\right|_t = \exp\left\{\frac{1}{2}\sigma^2 t + \sigma W_t\right\}.$$

Denoting the price of the quanto inverse call at t by C_t^{q}, its valuation under the new measure is trivial since it is a plain vanilla option and follows the exact same steps as the risk-neutral derivation of the Black-Scholes formula:

$$C_t^{\mathsf{q}} = \frac{B_t}{B_T}\, \mathbb{E}_t^{\mathbb{Q}}\left[\bar{X}\left(\frac{S_T - K}{S_T}\right)^+\right] = \mathbb{E}_t^{\tilde{\mathbb{Q}}}\left[\left(\left.\frac{d\tilde{\mathbb{Q}}}{d\mathbb{Q}}\right|_T\right)^{-1}\frac{\bar{X}}{S_T}\left(S_T - K\right)^+\right]$$

$$= e^{(\sigma^2 - 2r)\tau}\frac{\bar{X}}{S_t}\, \mathbb{E}_t^{\tilde{\mathbb{Q}}}\left[\left(S_T - K\right)^+\right] = e^{(\sigma^2 - 2r)\tau}\frac{\bar{X}}{S_t}\int_{-\infty}^{\infty}\left(S_T(z) - K\right)^+ \phi(z)dz, \qquad (4.13)$$

where $S_T(z) = S_t \exp\left\{\frac{1}{2}\sigma^2\tau + \sigma\sqrt{\tau}z\right\}$, z is drawn from a standard normal distribution, and $\phi(\cdot)$ is the corresponding probability density function. Note that

$$\left(S_T(z) - K\right)^+ = 0 \Leftrightarrow z \geq \frac{\ln\left(\frac{S_t}{K}\right) + \left(r - \frac{3}{2}\sigma^2\right)\tau}{\sigma\sqrt{\tau}} = d_3.$$

Thus, using $\Phi(\cdot)$ to denote the standard normal distribution function we can calculate the integral on the RHS of (4.13):

$$\int_{-\infty}^{d_3}\left(S_T(z) - K\right)\phi(z)dz = S_t\int_{-\infty}^{d_3}\exp\left\{\left(r - \frac{3}{2}\sigma^2\right)\tau + \sigma\sqrt{\tau}z\right\}\phi(z)dz - K\Phi(d_3).$$

Evaluating the integral yields the GBM price of the inverse call with strike $K^{\$}$ as:

$$C_t^{\mathsf{q}} = e^{-r\tau}\bar{X}\left[\Phi(d_2) - e^{(\sigma^2 - r)\tau}Y_t K\Phi(d_3)\right], \qquad (4.14)$$

where $d_2 = d_3 + \sigma\sqrt{\tau}$ is the same as we see in the Black-Scholes FX pricing formula, i.e. $d_2 = d_2^\$$ under the assumption $r^\mathbb{B} = 0$. A similar argument yields the time t GBM price of a quanto inverse put with strike K USD and maturity T as:

$$P_t^{\mathrm{q}} = e^{-r\tau}\bar{X}\left[e^{(\sigma^2 - r)\tau}Y_t K\Phi(-d_3) - \Phi(-d_2)\right]. \tag{4.15}$$

The two likelihood functions $\frac{d\mathbb{Q}^\mathbb{B}}{d\mathbb{Q}^\$}$ and $\frac{d\widetilde{\mathbb{Q}}}{d\mathbb{Q}}$, albeit very similar, result in different pricing and hedging properties for the two functions, driven by the difference in denomination between these two contracts. In the quanto inverse option case it is possible, and indeed more convenient, to start from the risk-neutral dynamics of the underlying because the payoff is denominated in USD. One may define the equivalent martingale measure for the quanto inverse of the underlying denominated in USD, and price the option accordingly. However, this approach is inappropriate for an inverse option when the payoff is denominated in BTC. Due to Siegel's exchange paradox, denomination conversion from USD to BTC should be performed first under $\mathbb{P}$. Then a risk-neutral measure in BTC can be established which is symmetrical to the risk-neutral measure in USD. This would be the appropriate measure for pricing derivatives denominated in BTC.

4.5.3 Properties

Figure 4.5 compares the USD-denominated inverse BS prices and payoffs (above) with the prices and payoffs for quanto inverse options (below). These are displayed above as a function of S_T, and we display prices for different maturities ranging from ten days to a year. As before, prices and strikes are denominated in USD and we deliberately omit the superscripts. The panels on the left show calls (blue) and on the right we have the puts (red). The inverse option payoffs have the familiar convex structure of the BS pricing function, with an increasing gamma and a positive vega, and the price approaches the payoff as the option approaches expiry. The longer the maturity, the more valuable the options, i.e. the theta is positive.

The quanto inverse option pricing functions depicted in the lower panels of Figure 4.5 behave very differently. The payoff to a quanto inverse call is capped above at $\bar{X}$. Deep ITM quanto inverse calls decrease in value as the time to maturity increases, where they are valued below their intrinsic values. Indeed, very deep ITM quanto inverse call prices could be much lower than one would think by simply looking at ATM option prices.[13] Furthermore, the convexity of the quanto inverse call pricing function changes as the underlying price increases: it starts as a convex function but then changes into a concave function as the option moneyness increases. Therefore the delta of a quanto inverse call is not a monotonic increasing function with respect to the underlying price, as it is for vanilla options. There

---

[13]The area around ATM and slightly ITM quanto inverse calls is extremely sensitive. At maturity, if the underlying price is 10% higher than the strike, the payoff would be roughly \$2050; if the underlying price is twice the strike the payoff will be \$11,250; and if the underlying price was 10 times the strike, the payoff would be \$20,250.

Figure 4.5: Inverse and Quanto Inverse Option Prices under GBM

Option payoffs and prices obtained using the Black-Scholes formula for the inverse options and our formula (4.13) for the quanto inverse options. Prices are represented as a function of the underlying price with a thicker line as the option approaches expiry. Time to maturities of 10 days, 3 months, 6 months and 1 year are shown. In these plots the first column displays calls in blue and the second displays puts in red; the upper row shows the inverse option prices and payoffs, and the lower shows the quanto inverse option prices and payoffs. All four plots are calculated using the same $K = \$25,000$, $r = 0\%$ and $\sigma = 75\%$ for all maturities but with different USD-denominated contingent claims (4.3) and (4.11), respectively. For the quantos we set $\bar{X} = \$22,500$.
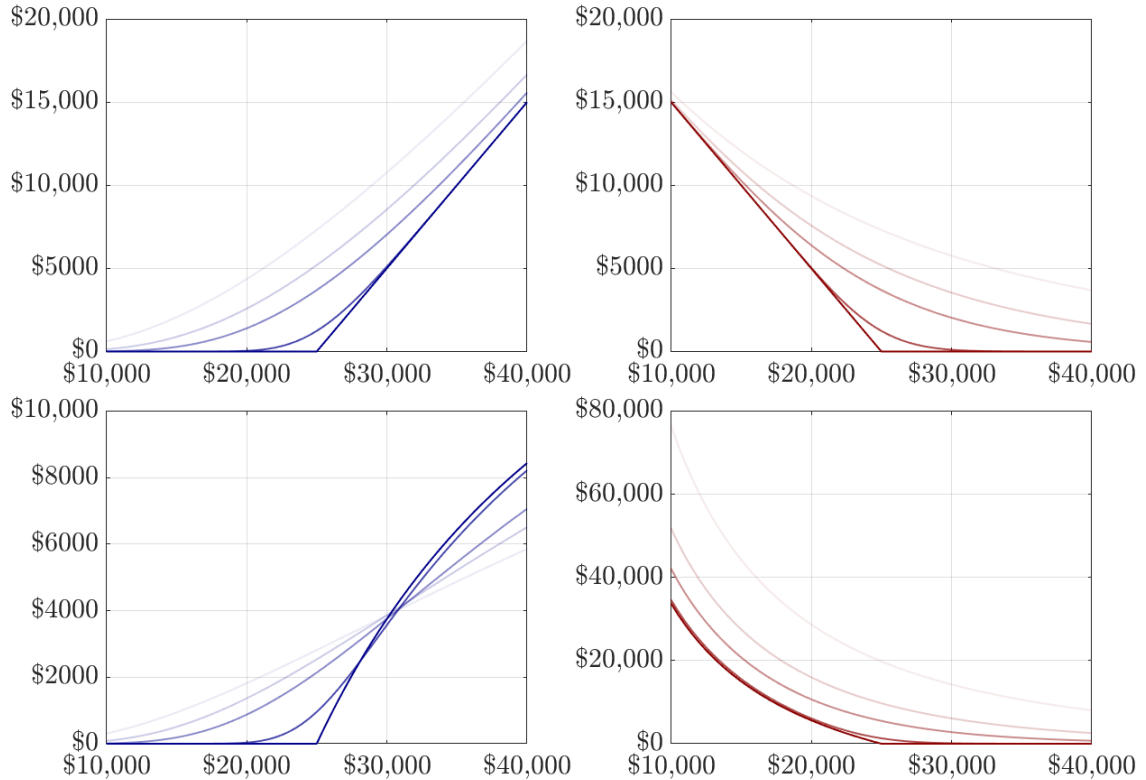
exists a global maximum at which gamma changes from being positive to being negative. For deep OTM and ATM options, the term structure of quanto inverse calls resembles that of inverse calls but for deep ITM options this pivots. Counter-intuitively, the price of a short-term deep ITM option exceeds the prices of their long-term counterparts, indicating a *negative* theta for these moneyness levels. The payoff to the quanto inverse put in Figure 4.5 is uncapped with the put price tending towards ∞ as the underlying price falls, whereas the payoff to a non-quanto inverse put converges to K. For all moneyness and maturities, the put pricing curves are convex decreasing with the strongest price sensitivity for ATM options. Unlike the inverse put delta, the quanto inverse put delta decreases monotonically with the underlying price, but it has no lower bound. The theta for a quanto inverse put is positive,

like its inverse option counterpart, i.e. the longer the time to maturity, the more valuable the option. It is interesting that the roles of calls and puts are now reversed, in that an inverse call and a quanto inverse put can – theoretically – pay an infinite amount to the holder, whereas the inverse put and quanto inverse call are capped at K and $\bar{X}$, respectively.

Now we investigate the quanto inverse option price dependence on the prefixed conversion factor $\bar{X}$. Figure 4.6 compares the maturity payoffs to inverse and quanto inverse options for two different values of $\bar{X}$. Suppose an agent who is optimistic, but not euphoric, about future returns enters a long position in a quanto inverse call. He has the greatest interest in negotiating a quanto factor as high as possible because this increases the slope of the terminal payoff. For instance, the bottom left graph in Figure 4.6 illustrates how the quanto inverse call payoff can exceed that of the inverse call when $\bar{X}$ is high. The advantage of this long position depends on the underlying having a maturity value between K and $\bar{X}$: if the underlying ends below the strike at maturity the option expires worthless; and above $\bar{X}$ the agent would profit more by entering a direct (or standard) option. Market makers could earn a premium by offering these two alternatives to traders seeking to profit from high future volatility. Now consider the inverse vs quanto inverse put, on the right in Figure 4.6. Here it is the writer not the buyer of the option who can use quanto inverses to their advantage. Assuming the quanto inverse put ends up ITM with $\bar{X} << K$ (top right plot), then the sell side would reduce their losses up to $\bar{X}$ but exponentially increase them afterwards. The premium on the quanto inverse options could be relatively low to write a call, but would need to be exceptionally high to write a put because, in theory, there is a non-zero probability of the asset price reaching zero resulting in an infinite loss for the writer of quanto inverse puts. Particularly for bitcoin this might be an attractive feature. Given the high average ATM prices seen in Table 3.2, retail investors might prefer a quanto inverse option in contrast to the plain vanilla inverse.

Figure 4.7 illustrates inverse and quanto inverse call prices as functions of the underlying, for a given fixed strike and interest rate, and for different levels of market volatility σ, and time to maturity. Inverse option prices display the familiar pattern of increasing with either time to maturity or volatility. But quanto inverse call prices can be decreasing with volatility (negative vega) as well as maturity (negative theta) as previously discussed. For instance, with a fixed volatility at 200%, the fair price of a 10-day quanto inverse option with an ITM strike level at $K = \$30,000$ is $4123, but this decreases as maturity increases to 90 days ($4080) and to 180 days ($3490). But the OTM option with strike $K = \$20,000$ is strictly increasing with maturity. The 90-day OTM quanto inverse option has positive vega whereas the ITM option has a vega which changes sign from positive to negative as the option moves deep ITM; the price increases from $4020 ($\sigma = 50\%$) to $4270 ($\sigma = 100\%$) and decreases afterwards to $4080 ($\sigma = 200\%$). The difference between ITM inverse and quanto inverse option prices is more pronounced than for OTM options, and it also increases with maturity and volatility. Especially for long-term options and/or during periods of high volatility, a quanto inverse call provides a very affordable alternative to a standard inverse call.

Figure 4.6: Payoff Sensitivity with Respect to $\bar{X}^{\mathbb{B}/\$}$

Payoff comparison of inverse and quanto inverse call (left column) and put (right column) with different predetermined exchange factor $\bar{X}$. The underlying at maturity ranges between $\$10,000$ and $\$40,000$ with a strike price at $K = \$25,000$. We distinguish between inverse (black) and quanto inverse (red) options and compare different $\bar{X}$ between each column, i.e. $\bar{X} = \$15,000$ (upper row) and $\bar{X} = \$35,000$ (lower row).
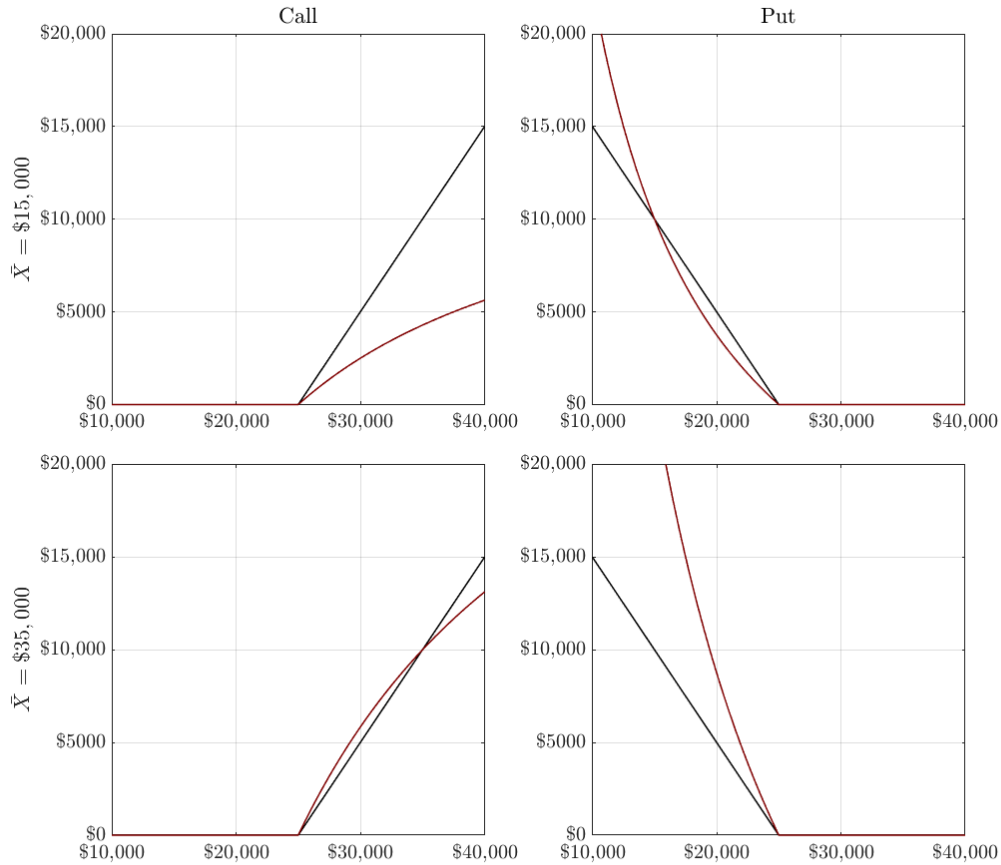
4.6 Payoff Summary

Table 4.1 summarizes the payoffs to different types of calls and puts with and without currency protection. As before, we use $\$$, $\mathbb{B}$ and $\mathbb{E}$ to denote payoffs in USD, bitcoin and ether, e.g. $\mathbb{B}1$ is one bitcoin or $\$10$ is ten dollars. Note that option strikes and underlyings are always quoted in USD. The upper panel considers standard and inverse options on the exchange rate BTCUSD with strike $K^{\$} = \$25,000$ and we report the payoffs for different settlement prices. The quanto versions suppose a pre-fixed exchange rate of $\bar{X}^{\$/\mathbb{B}} = \$22,500$. The lower panel is for options on the ETHUSD trading pair, where ETH is regarded as a security and as such could be replaced by any other crypto asset, as discussed above. In the table we let the option strike be $K^{\$} = \$1,750$ and again consider different scenarios for the settlement price. The lower part of Table 4.1 displays the paypoff to standard and inverse calls and puts on the ETHUSD trading pair. Below this the standard quanto option payoff is for a BTC-based trader in ETHUSD, seeking currency protection against a fall in the price of BTC relative to

Figure 4.7: Volatility-Maturity Price Sensitivity

Comparison of call prices as a function of the underlying using the inverse pricing formula (blue) and the quanto inverse pricing formula (red), given in USD. The figures depict different times to maturity (ranging from ten days to six months) under different constant underlying volatilities σ (ranging from 50% to 200%). We assume constant zero interest rate and fix both the underlying and $\bar{X}$ at \$25,000. We display only the 20% area around the strike price.

the dollar. The fourth payoff is for a USD-based trader entering a quanto inverse option on ETHUSD, seeking currency protection against a fall in the price of ETH relative to the US dollar. Notice the use of notation ETHUSD means we assume ETH is security here – and indeed could be replaced by any other crypto XYZ that is regarded as a security.

As already remarked, for every settlement price yielding a non-zero payoff, the standard call has a greater payoff than the quanto inverse call, but the standard put has a much smaller payoff than the quanto inverse put. This ordering holds for both BTCUSD and ETHUSD options, as shown in the table. Furthermore, the table compares the payoffs to other types of options. For instance, in the BTCUSD case (upper panel) the inverse and standard quanto options are both denominated in BTC, and again the calls and puts have opposite ordering. That is, the inverse call has a smaller payoff than the standard quanto but the inverse put has a greater payoff than the standard quanto. Finally we note that Table 4.1 only considers standard and inverse options because the direct options are similar to the standard ones, except the USD payoff is in a stablecoin.

Table 4.1: Payoffs to Standard and Inverse Options ± Currency Protection

We compare the payoffs to standard and inverse calls and puts with and without the currency protection provided by a fixed quanto factor. The currency of each payoff is recorded and numbers are rounded. The upper part is for options on BTCUSD, i.e. with $S^\$ =$ BTCUSD, with strike $K^\$ = \$25,000$ and different settlement prices from \$10,000 to \$50,000. The quanto factor for converting the standard option payoff from USD to BTC is $\bar{X}^{\$/\mathbb{B}} = \$22,500$, and for converting the inverse option payoff from BTC to USD it is $\bar{X}^{\mathbb{B}/\$} = \$22,500^{-1} = \mathbb{B}0.0000444$. The lower part is for options on ETHUSD, i.e. where $S^\$ =$ ETHUSD, with strike $K^\$ = \$1,750$ and different settlement prices from \$500 to \$2500. The quanto factor for converting the standard option payoff from USD to BTC is again $\bar{X}^{\$/\mathbb{B}} = \$22,500$, but now for converting the inverse option payoff from ETH to USD we use $\bar{X}^{\mathbb{E}/\$} = \$2,000^{-1} = \mathbb{E}0.0005$.

	payoff Function	$S_T^\$$				
Call		$10,000	$20,000	$30,000	$40,000	$50,000
Standard	$\left(S_T^\$ - K^\$\right)^+$	$0	$0	$5,000	$15,000	$25,000
Inverse	$\left(S_T^\$ - K^\$\right)^+/S_T^\$$	$\mathbb{B}0$	$\mathbb{B}0$	$\mathbb{B}0.17$	$\mathbb{B}0.38$	$\mathbb{B}0.5$
Standard Quanto	$\bar{X}^{\$/\mathbb{B}}\left(S_T^\$ - K^\$\right)^+$	$\mathbb{B}0$	$\mathbb{B}0$	$\mathbb{B}0.22$	$\mathbb{B}0.67$	$\mathbb{B}1.11$
Quanto Inverse	$\bar{X}^{\mathbb{B}/\$}\left(S_T^\$ - K^\$\right)^+/S_T^\$$	$0	$0	$3,750	$8,438	$11,250
Put						
Standard	$\left(K^\$ - S_T^\$\right)^+$	$15,000	$5,000	$0	$0	$0
Inverse	$\left(K^\$ - S_T^\$\right)^+/S_T^\$$	$\mathbb{B}1.5$	$\mathbb{B}0.25$	$\mathbb{B}0$	$\mathbb{B}0$	$\mathbb{B}0$
Standard Quanto	$\bar{X}^{\$/\mathbb{B}}\left(K^\$ - S_T^\$\right)^+$	$\mathbb{B}0.67$	$\mathbb{B}0.22$	$\mathbb{B}0$	$\mathbb{B}0$	$\mathbb{B}0$
Quanto Inverse	$\bar{X}^{\mathbb{B}/\$}\left(K^\$ - S_T^\$\right)^+/S_T^\$$	$33,750	$5,625	$0	$0	$0

	payoff Function	$S_T^\$$				
Call		$500	$1000	$1500	$2000	$2500
Standard	$\left(S_T^\$ - K^\$\right)^+$	$0	$0	$0	$250	$750
Inverse	$\left(S_T^\$ - K^\$\right)^+/S_T^\$$	$\mathbb{E}0$	$\mathbb{E}0$	$\mathbb{E}0$	$\mathbb{E}0.125$	$\mathbb{E}0.3$
Standard Quanto	$\bar{X}^{\$/\mathbb{B}}\left(S_T^\$ - K^\$\right)^+$	$\mathbb{B}0$	$\mathbb{B}0$	$\mathbb{B}0$	$\mathbb{B}0.01$	$\mathbb{B}0.03$
Quanto Inverse	$\bar{X}^{\mathbb{E}/\$}\left(S_T^\$ - K^\$\right)^+/S_T^\$$	$0	$0	$0	$250	$600
Put						
Standard	$\left(K^\$ - S_T^\$\right)^+$	$1,250	$750	$250	$0	$0
Inverse	$\left(K^\$ - S_T^\$\right)^+/S_T^\$$	$\mathbb{E}2.5$	$\mathbb{E}0.75$	$\mathbb{E}0.17$	$\mathbb{E}0$	$\mathbb{E}0$
Standard Quanto	$\bar{X}^{\$/\mathbb{B}}\left(K^\$ - S_T^\$\right)^+$	$\mathbb{B}0.06$	$\mathbb{B}0.03$	$\mathbb{B}0.01$	$\mathbb{B}0$	$\mathbb{B}0$
Quanto Inverse	$\bar{X}^{\mathbb{E}/\$}\left(K^\$ - S_T^\$\right)^+/S_T^\$$	$5,000	$1,500	$333	$0	$0

4.7 Hedge Ratios

Table 4.2 presents the pricing formulae for inverse and quanto inverse options and their hedge ratios, under the GBM assumption. As in section 4.3 we assume $r^{\mathbb{B}} = 0$ and set $r^{\$} = r$ for brevity, this way we have the same interest rate for both inverse and quanto inverse options, and for the sake if clarity we omit the superscript of the strike K. Hence, we omit the interest rate sensitivity ρ. We provide the mathematical derivation of the Greeks in the Appendix A.

Table 4.2: Inverse and Quanto Inverse Option Prices and Greeks

We assume the tradable underlying price S follows a GBM with volatility σ, where the drift depends on the USD discount rate r. Furthermore, we assume a zero BTC discount rate. The inverse or quanto inverse call or put have strike K and for the sake if clarity we omit the pre-defined $\bar{X}$ for the quanto inverse case. Note that this need to be multiplied to the individual Greek. The residual time to maturity τ hold for all types and we use the notation $\omega = \pm 1$ according as the option is a call or a put. Suppressing all time subscripts for simplicity we set $d_1 = \ln\left(\frac{S}{K}\right)[\sigma\sqrt{\tau}]^{-1} + (r + \frac{1}{2}\sigma)\sqrt{\tau}$, $d_2 = d_1 - \sigma\sqrt{\tau}$ as well as $d_3 = d_2 - \sigma\sqrt{\tau}$.

Name		Inverse	Quanto inverse
Price	f	$\omega[S\Phi(\omega d_1) - e^{-r\tau}K\Phi(\omega d_2)]$	$\omega e^{-r\tau}[\Phi(\omega d_2) - e^{(\sigma^2-r)\tau}S^{-1}K\Phi(\omega d_3)]$
Delta	δ	$\frac{\partial f}{\partial S}$ $\quad \omega\Phi(\omega d_1)$	$\omega e^{(\sigma^2-2r)\tau}\frac{K}{S^2}\Phi(\omega d_3)$
Gamma	γ	$\frac{\partial^2 f}{\partial S^2}$ $\quad \frac{\phi(d_1)}{S\sigma\sqrt{\tau}}$	$e^{(\sigma^2-2r)\tau}\frac{K}{S^3}\left[\frac{\phi(d_3)}{\sigma\sqrt{\tau}} - \omega 2\Phi(\omega d_3)\right]$
Vega	ν	$\frac{\partial f}{\partial \sigma}$ $\quad S\phi(d_1)\sqrt{\tau}$	$e^{-r\tau}\left[\phi(d_2)\sqrt{\tau} - \omega e^{(\sigma^2-r)\tau}\sigma\tau\frac{2K}{S}\Phi(\omega d_3)\right]$
Volga	ν^o	$\frac{\partial^2 f}{\partial \sigma^2}$ $\quad \sqrt{\tau}\phi(d_1)S\frac{d_1 d_2}{\sigma}$	$e^{-r\tau}\sqrt{\tau}\phi(d_2)\frac{d_2 d_1}{\sigma} - \omega 2\frac{K}{S}\tau e^{(\sigma^2-2r)\tau}\times$ $\left[\Phi(\omega d_3)(1+2\sigma^2\tau) - \sigma\phi(d_3)\left(\frac{-d_1}{\sigma} - \sqrt{\tau}\right)\right]$
Vanna	ν^a	$\frac{\partial^2 f}{\partial S \partial \sigma}$ $\quad -\phi(d_1)\frac{d_2}{\sigma}$	$\omega e^{(\sigma^2-2r)\tau}\times$ $\frac{K}{S^2}\left[2\tau\sigma\Phi(\omega d_3) + \omega\phi(d_3)\left(\frac{-d_1}{\sigma} - \sqrt{\tau}\right)\right]$
Theta	ϑ	$-\frac{\partial f}{\partial \tau}$ $\quad \omega[-re^{-r\tau}K\Phi(\omega d_2)]$ $-\frac{\sigma S\phi(d_1)}{2\sqrt{\tau}}$	$-e^{-r\tau}\left(\frac{\phi(d_2)\sigma}{2\sqrt{\tau}} - \omega r\Phi(\omega d_2)\right)$ $+\omega e^{(\sigma^2-2r)\tau}\frac{K}{S}(\sigma^2-2r)\Phi(\omega d_3)$

Figures 4.8 (4.10) and 4.9 (4.11) illustrate the delta, gamma, vega, theta, volga and vanna of inverse and quanto inverse calls (puts) as a function of K for fixed S, σ and r and for different τ. We set $\sigma = 75\%$ and we compare the inverse options hedge ratios in blue with the corresponding ratios for the quanto inverse option in red, for options of different maturities. First compare the deltas. The inverse call (put) delta has the usual shape of a monotonically increasing (decreasing) normal distribution function. But the quanto inverse call delta is not monotonic but has a maximum when the underlying price just exceeds the strike and declines thereafter eventually becoming zero for deep ITM options. This is because the payoff is capped at $\bar{X}$, i.e. even for large price movements the change in the payoff is limited, as has

already been discussed above when commenting on the shift from convexity to concavity in the pricing function. Comparing the term structure in the two call deltas, short-term OTM (ITM) inverse call deltas are smaller (larger) than their long-term counterparts. The quanto inverse call deltas are similar but invert this relationship.

Very frequent rebalancing of a delta-hedged position can have an adverse impact on the volatility of the underlying, especially for short-dated ITM standard calls or puts where the delta is close to ± 1 (Golez and Jackwerth, 2012; Ni et al., 2021). Now, the delta is so much greater on ITM direct (or standard) calls than it is for quanto inverse calls of the same moneyness, it even approaches zero for very deep ITM quanto inverse calls. Therefore the unwanted volatility impact of frequent rebalancing on a delta-hedged position will be very much less for quanto inverse calls. By contrast with deep ITM quanto inverse calls, a deep ITM quanto inverse put delta can be very much greater than the delta of a direct (or standard) put of the same moneyness, so the adverse volatility impact of rapid delta-hedge rebalancing referred to above would be exacerbated. And a delta-hedge of a short position on such an option could require buying more of the underlying than the option's notional. However, if this buying pressure causes the underlying price to rise the put would becomes less ITM and its delta would then decline. Finally, we note that long-term quanto inverse put deltas are generally lower than their short-term counterparts, irrespective of moneyness, except for short- and mid-term near-ATM options.

The inverse call gamma and vega are both positive, following the standard normal density and are identical for call and puts. But this is not the case for quanto inverse options because the non-monotonicity in the quanto inverse call delta influences the shape of its gamma. Specifically, quanto inverse call gammas have a similar shape to inverse call gamma at high strikes but the gamma becomes negative at lower strikes, before eventually converging to zero as the strike tends to zero. This is due to the change from a convex to concave pricing curve. Both inverse and quanto inverse put deltas are monotonic, but the quanto inverse deltas decrease faster as a function of strike, so for strikes above ATM the quanto inverse gamma is the greater of the two. Just as the standard option gamma decreases as time to expiry increases, the quanto inverse call gamma gradually decreases with maturity, without a lower boundary. The consequences become severe for hedging, as to be delta-neutral the issuer would need to buy more and more units of the underlying, eventually exceeding the notional of the option. We have already discussed the negative theta for inverse and quanto inverse options. Figures 4.9 and 4.11 show that the only inverse options with a positive theta are in fact low strike quanto inverse calls. Another notable feature from these figures is that the volga for longer-maturity quanto inverse puts can very large and positive, and that the vanna can take either sign. For inverse calls and puts it is negative for low-strike options and positive for high-strike options. For low-strike quanto inverse calls and puts the vanna is negative but for high-strike calls it is positive and for most high-strike puts it is negative.

Figure 4.8: Inverse and Quanto Inverse Call Greeks I

Inverse and quanto inverse Greeks for calls as a function of the strike level K with fixed underlying S_t and $\bar{X} = S_t = \$25,000$, volatility $\sigma = 75\%$ and different times to maturity: from 10 days (bold), 30 days (dashed) and 90 days (dotted). The left, blue column represents the inverse Greeks, while the right, red column shows the quanto inverse Greeks, using the same vertical scale for ease of comparison.
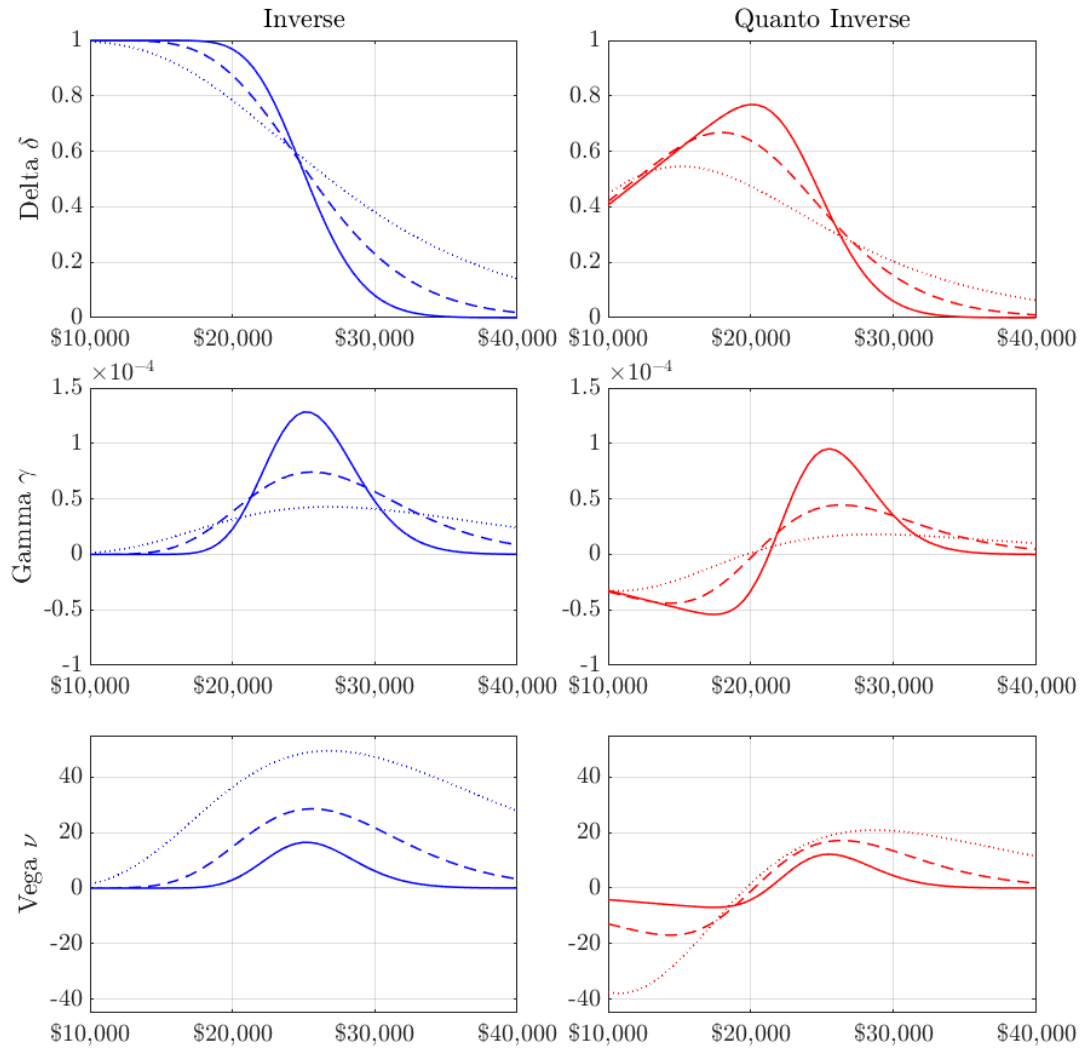
Figure 4.9: Inverse and Quanto Inverse Call Greeks II

Inverse and quanto inverse Greeks for calls as a function of the strike K with fixed underlying and predetermined conversion factor $S_t = \bar{X} = \$25,000$, volatility $\sigma = 75\%$ and different times to maturity: from 10 days (bold), 30 days (dashed) and 90 days (dotted). The left, blue column represents the inverse Greeks, while the right, red column displays the quanto inverse Greeks, using the same vertical scale for ease of comparison.
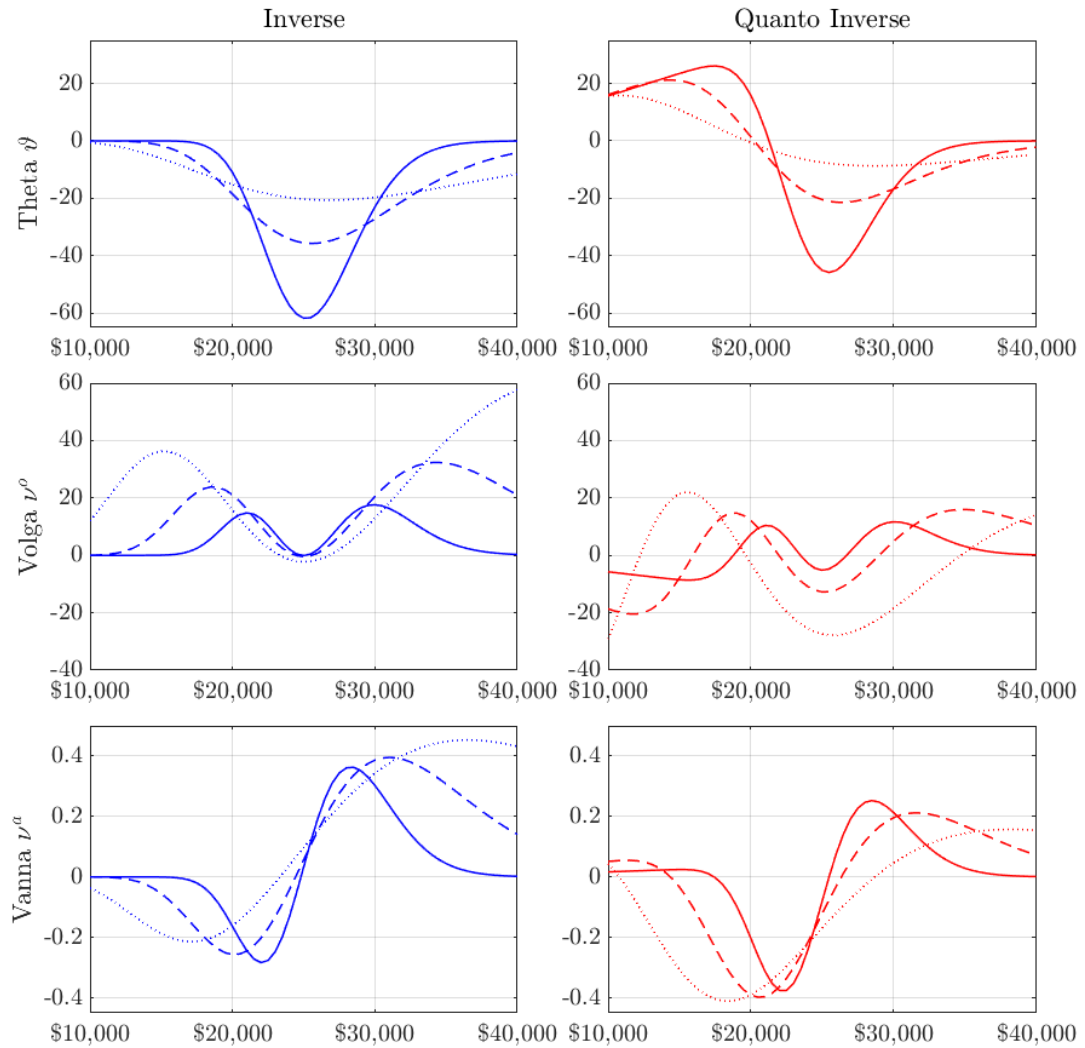
Figure 4.10: Inverse and Quanto Inverse Put Greeks I

Inverse and quanto inverse Greeks for puts as a function of the strike K with fixed underlying and predetermined conversion factor $S_t = \bar{X} = \$25,000$, volatility $\sigma = 75\%$ and different times to maturity: from 10 days (bold), 30 days (dashed) and 90 days (dotted). The left, blue column represents the inverse Greeks, while the right, red column displays the quanto inverse Greeks, using the same vertical scale for ease of comparison.
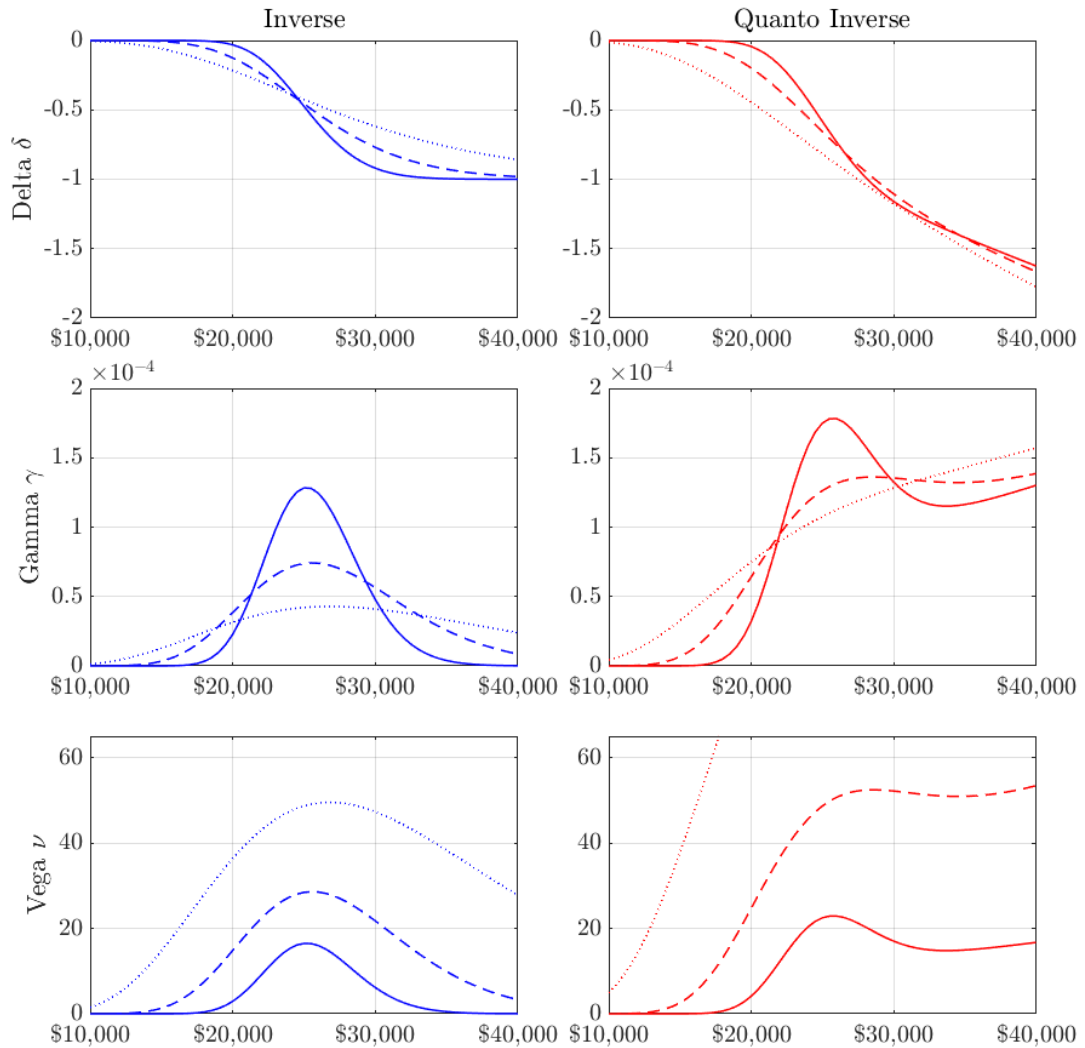
Figure 4.11: Inverse and Quanto Inverse Put Greeks II

Inverse and quanto inverse Greeks for puts as a function of the strike K with fixed underlying and predetermined conversion factor $S_t = \bar{X} = \$25,000$, volatility $\sigma = 75\%$ and different times to maturity: from 10 days (bold), 30 days (dashed) and 90 days (dotted). The left, blue column represents the inverse Greeks, while the right, red column displays the quanto inverse Greeks, using the same vertical scale for ease of comparison.
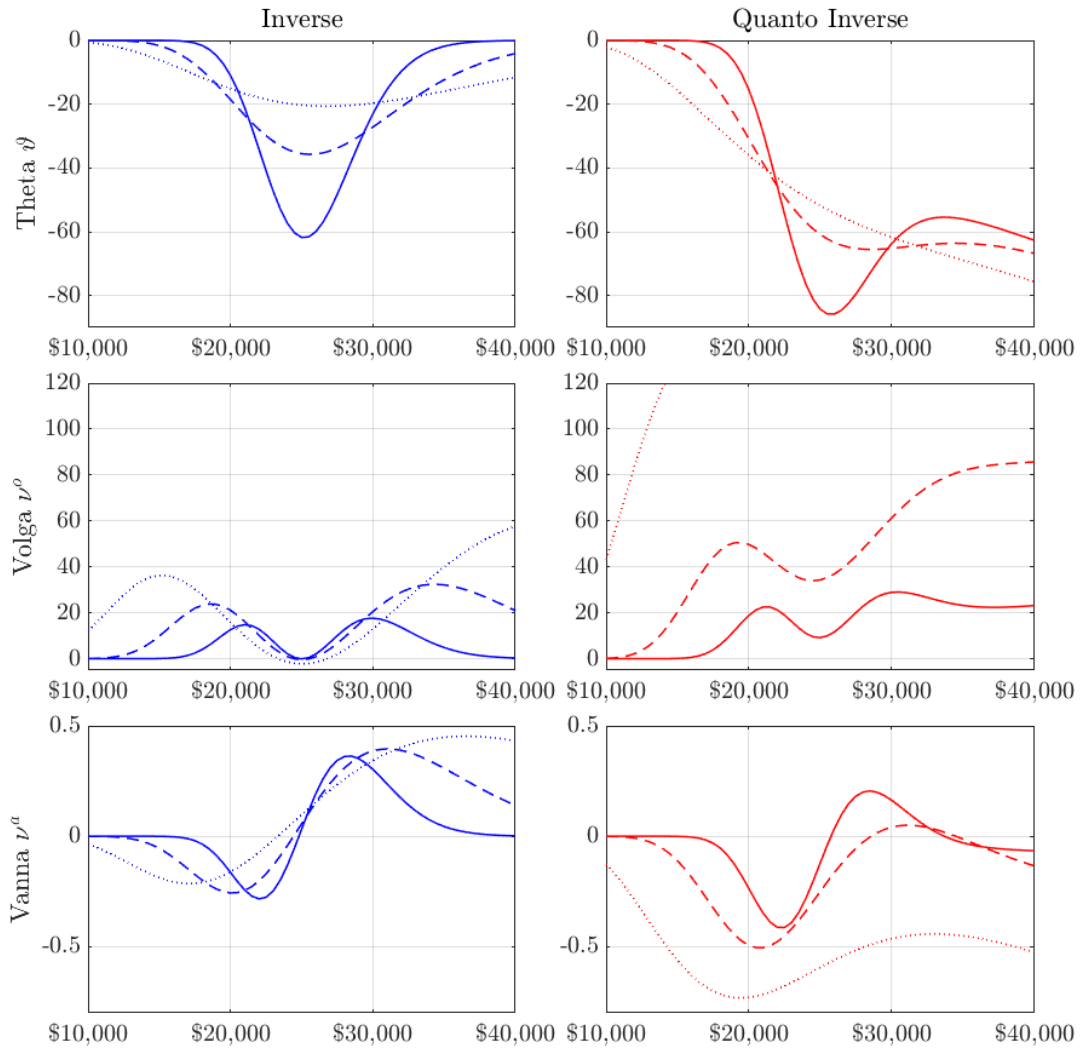
4.8 Concluding Remarks

Understanding the different products in the bitcoin option market is fundamental for any participants, not only for trading and hedging, but also concerning the exact payoff of their positions. In this chapter, we highlight the key differences between standard (or direct) and inverse options. While standard or direct options have a plain vanilla payoff and hence should can be prices using standard (stochastic volatility) option pricing models, we show that inverse options, which account for over 85% of trading on crypto options, have the same payoff structure as a standard FX option, and should therefore be priced in accordance to (Garman and Kohlhagen, 1983). Furthermore, we emphasise the mathematical models to price and hedge these, even if one side of the trading pair is regarded as a security.

Nevertheless, the Deribit options market, which virtually monopolises the trading on inverse options is theoretically incomplete. We indicate that accounting for this incompleteness requires a preference-based approach, under which the option price would depend on the trader's utility and risk tolerance. All Deribit options, and those on other unregulated exchanges are accounted in bitcoin, ether or a stablecoin, and only these currencies can be used for deposits and withdrawals. But it takes time for USD-based traders to convert between USD and these crypto, perhaps using a US-based exchange like Coinbase or Kraken. This friction induces non-zero liquidity premium on the prices of Deribit inverse options or any other crypto/stablecoin-denominated option, as investors are forces to transfer funds between exchanges and pay transaction costs.

In an attempt to mitigate these risks, we explain how quanto direct options can offer all traders protection against decoupling of a stablecoin from its 1:1 USD peg. Another use case for quanto crypto options is to provide exposure to another crypto without traders needing to change their base currency. As a natural extension, quanto inverse options pose a new type of exotic option which allow USD-denominated traders to gain exposure to the expanding crypto market without taking any crypto on the balance sheet. For instance, a trader could fix a BTCUSD quanto factor, so that any options that are settled in BTC have all profit and loss automatically converted to USD. The concave (call) and convex (put) payoff structure of quanto inverse options possess features that are attractive to both buyers and sellers. The uncapped USD put payoff is a perfect insurance against crypto price crashes. Motivated by the discussion and ATM prices of Section 3.2, we emphasise that the capped USD payoff for a quanto inverse call results in prices that are lower than standard (or direct) calls of the same moneyness and maturity, depending on the conversion factor and thus attractive for retail trader. Market maker can collect a premium by offering these production in addition to standard/direct or inverse options across exchanges.

CHAPTER 5

DELTA HEDGING BITCOIN OPTION

Motivation: Delta-hedging an options portfolio is a fundamental strategy that every options trader must employ to effectively manage their risk exposure. The Black-Scholes model, which is widely considered a benchmark for delta-hedging analyses, has been challenged as various empirical studies in financial literature have questioned its assumptions. Constant volatility in both traditional and crypto markets is not given, with implications on the delta-hedging position. Traders need to consider the correlation between the underlying and implied volatility to hedge effectively. Furthermore, exchanges that operate continuously, without closing, present opportunities to hedge positions continuously, thereby minimising the hedging error.

Summary: We conduct a comprehensive delta-hedging study for options with maturities up to one month and strike levels 30% above and below the underlying asset, using the Black-Scholes delta as a benchmark. Our analysis considers both fixed-maturity and perpetual futures as respective hedging instruments. The model-free deltas considered are smile-adjusted, smile-implied, and are meant to minimise the instantaneous variance of a delta-hedged portfolio. Given the 24/7 operational hours of the bitcoin options markets, our research does not restrict itself to daily rebalancing but instead adjusts positions every eight hours. While we are unable to identify a single model that outperforms the BS delta across all strike levels based on the variance of hedging errors, our findings do indicate promising results for ATM and deep OTM tails.

5.1 Delta Hedge

The assumption that implied volatility serves as an unbiased estimate of future realised volatility is inaccurate. In fact, the fair valuation of implied volatility should be slightly above the expected realised volatility, primarily due to the associated hedging costs. Hence, the practice of selling options and delta-hedging the position has become a popular trading strategy for equities and other asset classes to generate additional yield.[1] Nevertheless, this does not apply for cryptocurrency markets. Short position in equity volatility is similar to a long position the underlying equity, mainly due to the leverage effect.[2] We will demonstrate in this chapter that the correlation between the bitcoin price and its implied volatility is subject to frequent fluctuations and prolonged periods of positive correlation, making this form of yield farming untenable. Hence, understanding bitcoin's delta-hedging behaviour is of utmost importance.

Another key distinction between crypto and equity delta-hedging lies in the derivatives exchange structure. The Black and Scholes (1973) framework assumes, among other criteria, constant volatility and allows the delta position to be hedged continuously. However, these assumptions are not present in traditional markets because (i) volatility is not constant; and (ii), markets close overnight, weekends and bank holidays. Consequently, the payout of a delta-hedged option in traditional markets is the sum of the variance attributable to hedging with unknown volatility and the variance resulting from discrete delta-hedging. In certain aspects, cryptocurrency derivatives markets exhibit a more advanced state compared to their conventional counterparts. The majority of cryptocurrency exchanges facilitate trading on a continuous basis, 24 hours a day, 365 days a year, thereby permitting, at least theoretically, the continuous delta-hedging of a position. This increased frequency of hedging, even beyond standard trading hours, might minimise hedging errors. More significantly, it avoids potential price jumps between closing and opening times, as these are non-existent for crypto.

The assumption of constant volatility, and consequently zero correlation between the underlying price and its volatility has far-reaching impact on the hedging itself. While the benchmark for any study of dynamic delta-hedging is the Black and Scholes (1973) model, is well known that equity index options have a large and negative spot-volatility correlation which leads to a pronounced skew in the implied volatility curve. The sticky (local volatility) models, first introduced by Derman (1999), have become the industry standard of implied volatility-adjusted delta-hedging and we shall discuss them in greater detail in the next section. Following the basic idea from Bates (2005), and the more general results of Alexander

---

[1]The term 'delta' describes the sensitivity exhibited by an option with respect to changes in the underlying. The process of neutralising this exposure is referred to as 'delta-hedging'. This strategy consists of a long (short) position in the option and an opposing position in the underlying with size delta. Typically, the hedging is done with forwards of identical maturity as the option. However, cryptocurrency markets offer a suitable alternative in form of non-expiring perpetual swaps. This approach results in a convex payoff, thereby representing the most rudimentary form of volatility trading. Traders can capitalise on the discrepancies in implied volatility by continuously re-hedging their position, i.e. adjust size of the underlying position based on the changed of delta.

[2]The leverage effect here refers to the negative relationship between underlying return and implied volatility – see for example Bouchaud et al. (2001).

and Nogueira (2007a), it is possible to use the slope of the implied volatility curve to imply an adjustment to the BS delta which is model-free, in the sense that it is the same for any scale-invariant model. Indeed, Alexander and Nogueira (2007b) show that every stochastic and/or local volatility equity option pricing model for a tradable instrument falls into the scale-invariant class, as discussed already in Chapter 2. Consequently, the additional information gained from using any SV model become virtually negligible and differences arise only due to calibration error.

It is standard practice for equity option market makers to hedge their exposures using simple model-free adjustments to the BS delta, because these are regarded as so-called 'robust finance', i.e. the hedge ratios are model independent. The smile-implied and other smile-adjusted adjusted delta-hedges are particularly popular with practitioners, as evidenced by numerous articles and forums.[3] There are several previous empirical studies of smile-implied and/or smile-adjusted delta-hedging, but all of them study equity index options. Not all of the results are consistent, as seen in Chapter 2. Much less is known about the success of smile-adjusted delta-hedges for other types of options. The purpose of this chapter is to examine the performance of various smile-implied and other smile-adjusted delta-hedges when applied to bitcoin options. A great practical advantage of our study is that all deltas are very easy to compute. There is no requirement for model calibrations because all information is derived from the volatility smile in a straightforward and robust, model-free manner. We present results for delta-hedging using various adjustments of the BS delta which depend on the current regime of the market, the shape of the IV smile, and/or the spot-vol correlation.

Our focus is on short-term options with expiry ranging from ten to 30 days, which account for roughly 80% of all trading volume on bitcoin options as discussed in Chapter 3. Moreover, we need a proper smile to apply a smile-adjustment to the BS delta, and the range of liquid strikes for these short-term options is considerable. In fact, the moneyness of the options used in our empirical analysis ranges from 0.7 to 1.3. We only study dynamic delta-hedging with regular rebalancing, every eight hours at funding payment times or once per day at 00:00 UTC. This choice of experimental design is based on the bitcoin option market characteristics which are novel as explained in Chapter 1. The transaction costs for futures are very much smaller than they are for options. For instance, the spread on a futures contract ranges from approximately one to five basis points, depending on the expiry, but the spread on the short-term at-the-money options that would normally be used for gamma hedging are typically about 200 to 300 basis point. So gamma hedging is very much more expensive than regular dynamic delta-hedging. The transaction costs from rebalancing a gamma hedge could erode any profits made from reducing the hedging error, whereas the transaction costs from rebalancing a delta-hedge are tiny, especially when the perpetual contract is used as hedging instrument.

---

[3]See for instance, this recent CAIA article, another one on medium, and several quantitative finance forums such as risklatte and stackexchange.

5.2 Bitcoin Implied Volatility

The dynamic relationship between bitcoin's price and its implied volatility offers an intriguing point of investigation. The upper plot of Figure 5.1 illustrates the daily bitcoin implied volatility for ATM options with constant 30-day maturity, in red, interpolated from market data in an arbitrage-free way and the underlying bitcoin spot price in black from January 2020 to January 2023. At first glance, the implied volatility resembles other asst classes' implied volatilities with sudden upward jumps, followed by mean-reverting corrections. For instance, the red graph illustrates a sharp increase in implied volatility following abrupt price shocks, which can be attributed to diverse causes such as the COVID-19 pandemic or events related to digital platforms like and FTX. Implied volatility on such occasions has been observed to reach peaks of up to 170%, but subsequently, adjusting itself in a mean-reverting manner towards an average level approximating 75%. During calm, range-bounded periods like the last quarter of 2022, the implied volatility tends to decrease further, reaching lowest levels of just below 50%. Interestingly, we find an extensive period around January 2021 where the implied volatility is slowly building up, rather than explosively jumping. Additionally, the correlation between spot and volatility for bitcoin exhibits greater complexity than what is typically observed for equity indices or commodities. A pertinent example is the quarter leading up to January 2021, during which implied volatility appears to increase concurrently with an increase in the underlying value. Additionally, we find an interesting relationship between these variables in the period spanning August 2021 to May 2022, where there appears to be very low correlation.

For a deeper analysis of the spot-volatility relationship, the lower graph of Figure 5.1, presents the correlation between bitcoin price returns and 30-day constant maturity ATM implied volatility returns, depicted daily between January 2020 and January 2023. We find frequent changes, not only in size but also in the sign if the correlation. This oscillation, particularly its (frequent) shifts towards the positive territory, is somewhat atypical when compared with traditional markets, where the spot-volatility correlation is negative. Notably, in the period preceding the COVID-19 crash of March 2020, we observe a positive correlation that approaches 0.6. This correlation then pivots into the negative domain, falling to approximately -0.7 following the crash, before subsequently stabilising and fluctuating around zero over the next six months. While equity trader might hedge a long position on S&P 500 volatility by a short position in the index, this position might work adversely for crypto and point towards the same direction. Overall, we identify periods of high positive correlation (August and November 2020; September 2021; November 2022), high negative, as well as intervals of low correlation, ranging up to ±0.25 (June 2020; October 2021). This lends credibility to the hypothesis that the correlation between bitcoin returns and implied volatility is regime-dependent, with implications on not only the hedging of ATM, but also OTM options.

Figure 5.1: Bitcoin Spot, ATM Implied Volatility and its Correlation

Daily implied volatility of ATM bitcoin options with 30-day constant maturity (red; upper plot) in percentage points and the underlying bitcoin price in USD (black; upper plot). The sample covers the three-year period from January 2020 to January 2023. We display the forward looking 30-day correlation of daily implied volatility returns with bitcoin returns (blue; lower graph) over the same period on the same frequency.
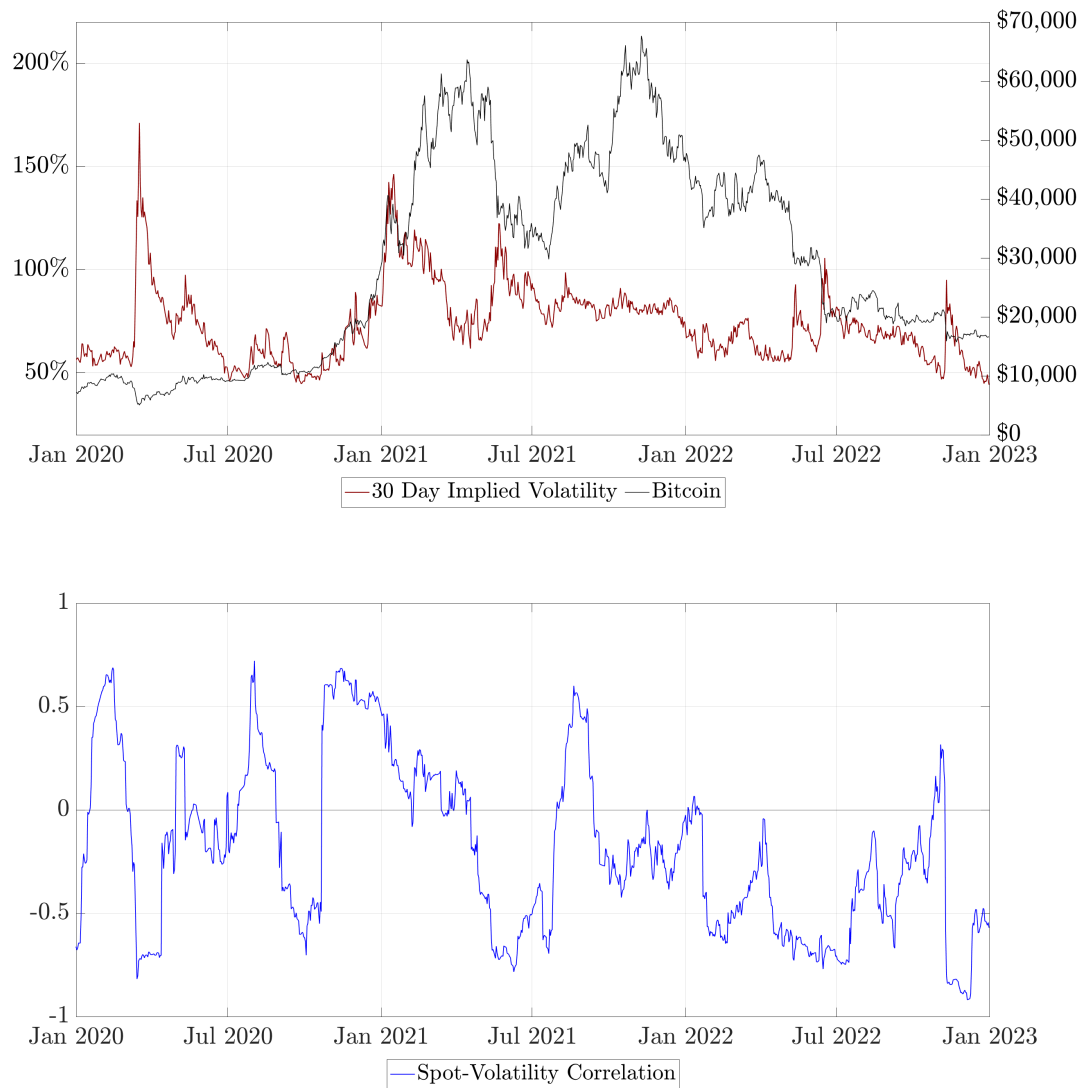
Figure 5.2 illustrates the dynamics of the implied volatility curve derived from Deribit options, charting its structure daily over a three year period.[4] Given the high liquidity for OTM options, we consider OTM puts for the moneyness 0.8 to 1, and OTM calls 1 to 1.2. The depicted data is interpolated in an arbitrage-free way at a fixed-maturity of 30-days throughout. The shape of the curve varies considerably over time. During the period around the 'Black Thursday' event in March 2020, when the bitcoin price fell over 30% in a few hours, the implied volatility curve was almost flat at the highest level recorded, i.e. around 170%. Typically, market makers add a high premium to options after sudden crashes due to the high uncertainty and panic in the market. Shortly after, the implied volatility developed the negative skew shape which is typical of equity index options, where OTM puts have much higher volatility than OTM calls. However, bitcoin options have very much higher implied volatilities than equity index options, in general.

During much of the sample period the implied volatility curve resembles a hockey stick, which at particularly tranquil times flattens into a slight symmetric smile. There are also instances of a positive skew, e.g. January 2021, where OTM calls have much higher volatility than OTM puts. The features exhibited are not commonly observed in equity index options markets, where the name 'skew' rather than 'smile' almost always applies. Throughout most of the sample, OTM puts with moneyness 0.8 have the highest implied volatility. In traditional (equity) markets, these deep OTM puts are an attractive means of insurance against falling prices. For instance, in the S&P 500 the pronounced and almost linear skew shape of the implied volatility curve means that the option prices that increase most following a fall in the underlying are those with lowest moneyness. By contrast, the bitcoin implied volatility curve was relatively symmetric prior to the crash on 12 March 2020. The lowest volatility of about 50% was for ATM options and both OTM puts and calls had roughly equal but higher volatilities, at around 75% for moneyness 0.8 and 1.2. However, a clear asymmetry in the smile emerges after price crashes, e.g. May 2021 or November 2022, with OTM puts drawing a higher premium from risk-averse investors seeking insurance against another significant price drop. Throughout 2022, there was a pronounced negative skew in bitcoin, but this shape is still much flatter than the skew one normally observes for equity index options. This asymmetry persisted but gradually diminished as the level of implied volatility dropped and the shape of the implied volatility curve once more began to resemble a smile as seen in December 2022.

The ATM implied volatility appears to be the lowest point of the smile with a negative skew during most of our sample data but, unlike equity index options, there are periods of high volatility, when the smile has a (strongly) pronounced positive skew. For instance, the slope of the smile increased during bitcoin's rally in June 2021, exhibiting a positive skew for several months. And whereas the equity index spot-volatility correlation is almost always large

---

[4]Implied volatility smiles emerge due to a variety of different factors. For example, put options are can be seen insurance, higher priced for sharp market falls, while call options serve as speculative tools. Divergent risk perceptions among market participants, skewed returns distributions, and the volatility-of-volatility contribute to these smiles further. Ultimately, supply and demand dynamics play a pivotal role.

and negative, the correlation between bitcoin and its implied volatility appears to be regime-dependent, as seen in Figure 5.1. However, it is important to note that other asset classes like FX, commodities or some single stocks present changing spot-vol correlation and alternating implied skews, see for example Carr and Wu (2007). Nevertheless, some features are not unlike those of equity index option implied volatilities: (i) volatilities at different moneyness move along with ATM volatility of the same maturity in a highly correlated fashion; and (ii), bitcoin's implied volatility term structure exhibits regular swings between high-volatility periods of backwardation and relatively calm periods of contango.[5]

Figure 5.2: Bitcoin Implied Volatility Curves and Term Structure

Implied volatility curves for bitcoin options with 30-day constant maturity, daily between 1 January 2020 and 30 June 2022, derived from OTM and ATM options. Implied volatilities are interpolated in an arbitrage-free way. Strike levels range from 20% below to 20% above the current value of the underlying BTC index.
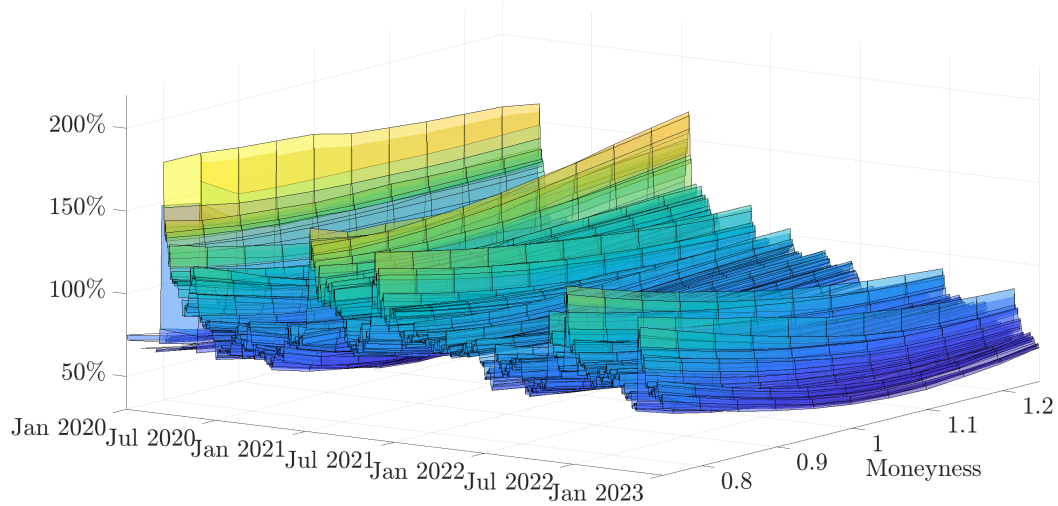
Figure 5.3 depicts the bitcoin ATM implied volatility term structure over the same period and frequency for 10, 20, 30, 60, 90 and 120 days constant maturity. It shows that, similar to equity index volatility term structures, bitcoin implied volatilities move along closely with little dispersion during most of the backwardation periods. In fact, throughout our three-year sample, we find only two occasions (July 2020, December 2022) in which the difference

---

[5]Eross et al. (2019) use 5-minute bitcoin spot data and find trading patterns which indicate a rise in volume during the day, but decline starting around 4 pm UTC and continuing until midnight. The authors conclude that this trend aligns with the typical intraday fluctuations seen in the FX markets. The bitcoin realised volatility remains relatively stable throughout the day but peaks around the opening hours of the world's three key stock markets, i.e. US, Europe and Asia. Finally, they find that bitcoin markets generally experience lower liquidity in the early morning hours. Wang et al. (2020) find similar results. The authors conclude that bitcoin spot market exhibit higher trading volume and volatility during European and US daytime hours. Shen et al. (2022) analyse the bitcoin intraday momentum and find significantly opportunities to predict intraday movements both in- and out-of-sample. Furthermore, the authors conclude that intraday momentum-based trading yields substantial economic gains and that that intraday momentum is appears stringer on high trading, high volatility days.

between short- and ultra long-term implied volatility is greater than 15%. On the other hand, during contango periods (March 2020, January 2021 or November 2022), the difference between long- and short-maturities can reach up to 85%, although these occasion appear to happen rather rarely and we see an immediate correction.

Figure 5.3: Bitcoin Implied Volatility Term Structure

Implied volatility term structure for bitcoin options with 10, 20, 30, 60, 90 and 120 days constant maturity, daily between 1 January 2020 and 1 January 2023, interpolated to constant from at-the-money strike levels in an arbitrage-free way.
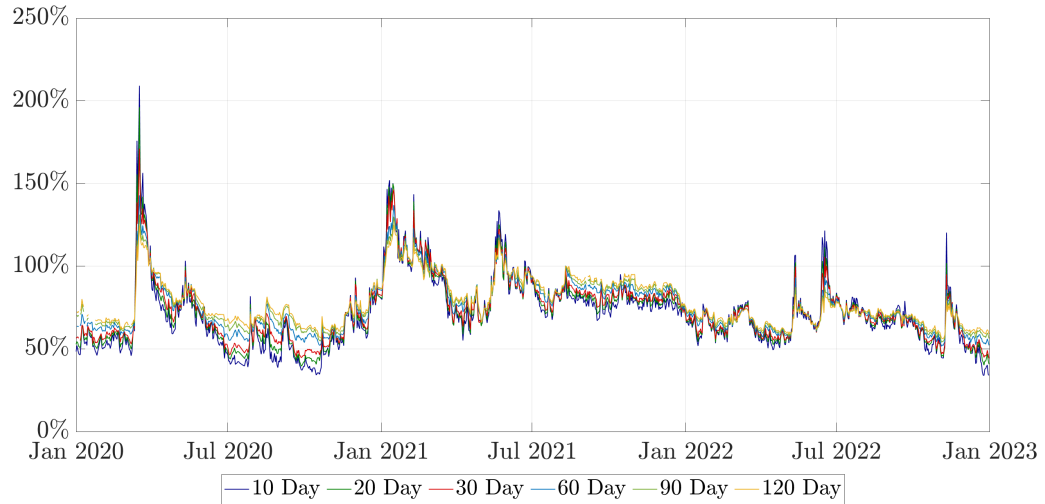
We conclude this section by using the characteristics of bitcoin options that we have highlighted above to motivate the rest of this chapter. A long bitcoin holder might buy one out-of-the-money put for insurance against a large price fall and consider the spot position suitably hedged. But market makers and other professional traders actively engage in dynamic delta-hedging because, as liquidity providers, it is essential for them to hedge the risk of writing options. They could do this using the BS delta, but in view of the popularity among equity option traders of smile-adjusted deltas it is interesting to study the effectiveness of such deltas for bitcoin options. We have reviewed a literature which debates the effectiveness of smile-adjusted deltas for hedging equity index options which reports many circumstances when the BS delta works just as well as any smile-adjusted one. However, no previous study has examined this question for bitcoin options, and it is clear – from the very different behaviour of the bitcoin implied volatility curve that we have just discussed as well as the array of new hedging instruments that are available for bitcoin – that one cannot simply extrapolate what is known about options on equity index options to reach a conclusion about hedging bitcoin options. Therefore, the goal of this study is to introduce and compare the various smile-adjusted deltas that are commonly used by practitioners to analyse their effectiveness for minimising the standard deviation of the hedging error for bitcoin options, based on different

choices of hedging instrument. In fact, the study could be extended even further, to the level of the exchange where the option is traded and/or hedged. For instance, is it better to use Binance or Deribit futures or perpetuals to hedge the options listed on the Deribit exchange? But we do not discuss this level of detail of the bitcoin option hedging problem in this study. At least currently, the Deribit options marketplace represents over 90% of trading volume on all bitcoin options at the time of writing, and personal correspondence with the Deribit option market makers indicates that they only use the Deribit futures platform for their delta-hedging activities.

5.3 Methodology

In our experimental design we write a standard European option on a Deribit bitcoin index futures worth one bitcoin and delta-hedge the position by entering a long position in a certain amount of a futures contract. The T-maturity futures allow traders to enter an agreement to buy or sell a certain amount of bitcoin at a future time T at a bitcoin-USD exchange rate agreed now. The underlying of both futures and options is the Deribit bitcoin index, BTC which is an aggregate non-tradable index. However, we could also hedge a T-maturity option with a position on the perpetual contract, instead of the T-maturity futures contract. We can suppress the running time t in our notation without confusion, and we denote the time t price of an inverse option with strike K and maturity T as $f(K, T \mid F, \sigma)$, where F is either the perpetual price or the T-maturity futures price, at time t, and $\sigma := \sigma_t(K, T|F)$ denotes the option's implied volatility also at time t.[6] By incorporating the relationship between volatility and the underlying in our hedging framework, we aim to achieve a more accurate delta than the BS delta, δ^{BS} using a smile-adjusted delta δ^{adj} which is based on the chain rule:

$$\delta^{adj}(K, T|F, \sigma) = \frac{\partial f}{\partial F} + \frac{\partial f}{\partial \sigma}\frac{\partial \sigma}{\partial F} = \delta^{BS}(K, T|F, \sigma) + \nu^{BS}(K, T|F, \sigma) \ \sigma_F(K, T|F, \sigma),$$

where δ^{BS} is the standard BS delta, ν_{BS} is the volatility-sensitivity of the BS option price (vega) and $\sigma_F = \frac{\partial \sigma}{\partial F}$ is the volatility-price sensitivity, i.e. the change in implied volatility for changes in the underlying. While the BS delta and vega have a closed-form formula and are easy to calculate, σ_F is rather difficult to quantify and several different approaches exist.

The first adjustments to the BS delta that we discuss have their roots in different parameterisations of local volatility depending on the current state of the market, or 'market regime'. Starting with classic papers by Dupire (1994) and Derman et al. (1996) the concept of local volatility has been developed in a broad-ranging academic literature. Of particular interest here are the 'sticky models', advocated by Derman (1999) for hedging equity index options, in the context of applying different parameterisations for the local volatility at the nodes in the binomial tree that models the underlying price evolution. Derman et al. (1996) proposed

---

[6]Note that we are transition form a C (call) and P (put) notation as used in Chapter 3. We are following standard literature convention.

an approximation to σ_F as the slope of the implied volatility with respect to the strike:[7]

$$\sigma_F(K,T|F) \approx k\ \sigma_K(K,T|F), \tag{5.1}$$

where $\sigma_K = \frac{\partial \sigma}{\partial K}$ is the derivative of the volatility with respect to the strike level and k should depend on the prevailing market regime. In fact, Derman (1999) introduced three different 'sticky models' to represent the behaviour of local volatility in different market regimes. The sticky strike model describes a trending market situation where he assumes the volatility to be independent of future price movements of the underlying and, as with the BS assumptions, it is constant and the same for every option. The delta in this regime is just equal to the BS delta. The sticky moneyness (also sometimes called the sticky delta) model considers a range-bounded market. During this regime an option's volatility depends only on its moneyness (or equivalently delta). So the local volatility is again the same at every node in the tree, but each option has a different tree, with a different local volatility, depending on the option's moneyness. As the underlying price moves, so the option's moneyness changes, we must move to a different tree to price the option.[8] Finally, the sticky tree model captures the local volatility behaviour in a rapidly falling market, i.e. it describes the smile adjustment when there is a strong negative correlation between volatility and the underlying price. This implied tree model gets its name from the local volatility model proposed by Derman and Kani (1994). Again, the local volatility is a deterministic function, but it can be different at each node in the tree, and the same tree is used to price all options. Under these three different types of parameterisation for local volatility, the values for k in (5.1) would differ according to the market regime, as follows:

$$k = \begin{cases} 0 & \text{(SS)} \\ 1 & \text{(ST)} \\ -K/F & \text{(SM)} \end{cases}.$$

Both Crépey (2004) and Alexander et al. (2012) extend the approximation (5.1) to add state-dependence to k. Also note that a little algebra, combining equations (1) and (2) of Alexander et al. (2012) with equation (3) of Alexander and Nogueira (2007b) shows that the smile-implied, scale-invariant delta of Bates (2005) is identical to the sticky-money approximation.

Considering that bitcoin is very volatile, the range of available strikes changes considerably over time. Therefore, to provide the framework for examining options with identical characteristics over a longer time horizon we therefore switch from a strike to a moneyness metric. We now denote the implied volatility by $\theta(m,T|F) = \sigma(mK,T|F)$. Denoting the partial derivatives of $\theta(m,T|F)$ with respect to F and m as $\theta_F(m,T|F)$ and $\theta_m(m,T|F)$

---

[7]Such an approximation has also been advocated by Coleman et al. (2001) and many other authors since.

[8]In practice, implied volatility is usually floored around the lowest ATM level and capped close to all-time highs of realised volatility, which leads to an range-bounded implied volatility over long time. The surface in absolute dimensions needs to shift to keep the surface in relative dimensions. This idea is referring to the sticky delta model.

respectively, we can rewrite the adjusted delta as:

$$\delta^{adj}\left(m,T|F\right) = \delta^{BS}\left(m,T|F\right) + \nu^{BS}\left(m,T|F\right)\ \theta_F\left(m,T|F\right),$$

and also rewrite (5.1) in the moneyness metric as:

$$\theta_F\left(m,T|F\right) \approx \kappa\,\theta_m\left(m,T|F\right). \tag{5.2}$$

We estimate the volatility-price sensitivity θ_F using the local volatility assumptions proposed by Derman (1999) where the type of tree used to model the option price evolution varies according to three possible market regimes: a stable-trending market (SS), a range-bounded market (SM) and a crash-jump regime (ST). This way, translating the sticky deltas of Derman (1999) into the moneyness metric, the values for κ in (5.2) should differ according to the market regime as follows:

$$\kappa = \begin{cases} 0 & \text{(SS)} \\ 1/F & \text{(ST)} \\ -m/F & \text{(SM)} \end{cases}.$$

As before, the model-free, smile-implied, scale-invariant delta of Bates (2005) and Alexander and Nogueira (2007a) is identical to the sticky moneyness delta of Derman and Kani (1994).

Next we consider the minimum variance (MV) delta δ^{mv}, i.e. the delta that minimises the instantaneous variance of a delta-hedged portfolio. Here we follow Bakshi et al. (1997) who introduce an approximation which minimises local variance. Lee (2001) shows that this MV hedge ratio has an adjustment of the same size as the (SM) smile-implied delta but with the opposite sign, that is:

$$\delta^{mv}\left(m,T|F\right) = \delta^{BS}\left(m,T|F\right) + \frac{m}{F}\ \nu^{BS}\left(K,T|F\right)\ \theta_m\left(m,T|F\right).$$

As explained in detail in Chapter 4 of Alexander (2008), and in other texts on implied volatility, the smile-implied delta produces a counter-intuitive 'floating smile' dynamic which also implies that the SM adjustment produces a hedging performance that is significantly worse than the BS delta when the volatility-price correlation is large and negative, i.e. when there is a pronounced negative skew. Since the MV adjustment has the opposite sign to the SM adjustment, the MV delta should out-perform the BS delta for hedging equity index options, and indeed for any option where the implied volatility curve has a marked negative slope.

Our final smile-adjusted delta which we denote δ^{hw} was proposed by Hull and White (2017). It is derived from an empirical estimate of a quadratic relationship between the absolute value of the daily PnL ΔP of a BS delta-hedged portfolio with value P, and the BS delta. That is:

$$y = \frac{\nu^{BS}}{\sqrt{\tau}}\frac{\Delta F}{F}\left(a + b\,\delta^{BS} + c\,(\delta^{BS})^2\right) + \epsilon.$$

where ΔF is the daily PnL of the futures. Having used historical data to obtain the parameter

estimates $(\hat{a}, \hat{b}, \hat{c})$, the Hull and White (HW) delta is then calculated as follows:

$$\delta^{hw} = \delta^{BS} + \frac{\nu^{BS}}{F\sqrt{\tau}}\left(\hat{a} + \hat{b}\,\delta^{BS} + \hat{c}\,(\delta^{BS})^2\right)$$

where δ^{BS} and ν_{BS} denote the classic BS delta and vega. The current underlying price is denoted as F, its change is denoted ΔF and τ is the option's time to expiry. We estimate $(\hat{a}, \hat{b}, \hat{c})$ through a least square regression over a rolling window adjustable size.

This section has covered a wide range of simple adjustments to the BS delta which which have proven effectiveness in previous studies of hedging options on stock indices and in other traditional asset classes. The question is now whether they can also out-perform a simple BS delta-hedge in bitcoin options markets – which are less mature than most traditional options markets, have more pronounced volatility and directional buying pressures, and where market makers rebalance their inventories based on information from these pressures. We summarise the BS-adjusted delta-hedge ratios considered in this study in a single formula as follows:

$$\delta^{adj} = \begin{cases} \delta^{BS} & \text{(SS/BS)} \\ \delta^{BS} + \nu^{BS}\theta_m \frac{1}{F} & \text{(ST)} \\ \delta^{BS} - \nu^{BS}\theta_m \frac{m}{F} & \text{(SM/SI)} \\ \delta^{BS} + \nu^{BS}\theta_m \frac{m}{F} & \text{(MV)} \\ \delta^{BS} + \nu^{BS}\frac{1}{F\sqrt{\tau}}\left(a + b\delta^{BS} + c\,(\delta^{BS})^2\right) & \text{(HW)} \end{cases} \qquad . \qquad (5.3)$$

We conclude with some remarks on the above:

1. The MV adjustment is identical to the ST adjustment when $m = 1$, i.e. for ATM options, otherwise it is greater in magnitude than the ST adjustment when $m > 1$, i.e. for OTM calls and smaller in magnitude than the ST adjustment when $m < 1$, i.e. for OTM puts;

2. The MV adjustment is always equal and opposite to the SM adjustment, and the SM delta is also the scale-invariant model-free delta of Alexander and Nogueira (2007a), i.e. the delta of any type of stochastic volatility jump process for the bitcoin option price;

3. The sign of the ST, SM and MV adjustments depends on the slope of the implied volatility curve, θ_m. When it has a negative slope the MV and ST deltas are less than – and the SM/SI delta is greater than – the BS/SS delta. When it has a positive slope the MV and ST deltas are greater than – and the SM/SI delta is less than – the BS/SS delta.

5.4 Empirical Results

Motivated by our discussions in Sections 3.2, 4.3 and 5.2 we regard the inverse option as a plain vanilla FX option, i.e. we convert its bitcoin price to the corresponding USD value using the current value of the option's underlying. Our focus is on short-term options with expiry ranging from 10 to 30 days, which are far more liquid and have a much wider strike range than the options studied by Matic et al. (2023).[9] This choice is because bitcoin options with maturity between one and three months represent only 20% of total trading volume and roughly 80% of all trading volume on bitcoin options is on options that expire in 30 days or less. Moreover, we need a proper smile to apply a smile-adjustment to the BS delta, and the range of liquid strikes for these short-term options is considerable. In fact, the moneyness of the options used in our empirical analysis ranges from 0.7 to 1.3. Our design is constructed for rebalancing the hedge either every eight-hours or every day, and the sample spans a two-year period from 1 January 2020 to 1 January 2022 which is divided into two one-year samples for presentation of the results. This is is mainly motivated by the traded volumes and differences in futures and perpetuals basis, as discussed in Chapter 3.

At each time t we short a European option with moneyness m and maturity T and hedge this with a long position in either the perpetual or the futures contract with the same maturity as the option and record the PnL of this portfolio as the hedging error under the physical measure, in the usual manner – see Hull and White (2017) for example.[10] Intraday market moves can be very considerable and the transactions costs from rebalancing are tiny, as already discussed. Therefore, we set the base frequency for presenting the tables of results to eight hours. We also time the eight-hourly rebalancing with the funding payment on the perpetual, i.e. at 00:00, 08:00 and 16:00 UTC. This is because the rebalancing of a hedge using the perpetual contract could be also be used to make profits from its funding payments.

We only study dynamic delta-hedging with regular rebalancing, every eight hours at funding payment times or once per day at 00:00 UTC. This choice of experimental design is based on the bitcoin option market characteristics which are novel as discussed earlier. The transaction costs for futures are very much smaller than they are for options. For instance, the spread on a futures contract ranges from approximately one to five basis points, depending on the expiry, but the spread on the short-term at-the-money options that would normally be used for gamma hedging are typically about 200 to 300 basis point. So gamma hedging is very much more expensive than regular dynamic delta-hedging. The transaction costs from rebalancing a gamma hedge could erode any profits made from reducing the hedging error, whereas the transaction costs from rebalancing a delta-hedge are tiny, especially when the

---

[9]Options used for the empirical study are interpolated in an arbitrage-free way. This provides a sense of hedging performance across a wide grid of maturity and moneyness. Through interpolation to constant maturity/strike, we are able to compare the models directly and get deeper insight into the hedging behaviour.

[10]Note that the option's delta differs whether one uses the futures or perpetuals contract. We remain the delta derived from the underlying as Deribit defines it and use both futures and perpetuals as (proxy-)hedge based on the calculated deltas.

perpetual contract is used as hedging instrument. In fact, these two one-year periods are already sufficient to to highlight key properties of the delta-adjustments at hand. Any further period would add value only in terms of robustness – which we already include – rather than extending findings. Furthermore, as seen in Figure 3.2 the basis of the futures/perpetuals during the last year of our sample is virtually negligible and hence would not yield any additional information. Data from August 2019 is indeed available but the trading volumes in the second half of 2019 were too low, and traded strike levels too narrow to consider useful.

All the deltas in (5.3) except the HW delta require us to calculate the slope of the implied volatility curve every time we rebalance the hedged portfolio. We examined various numerical techniques to calculate the derivative of the implied volatility curve, finding that fitting a polynomial of degree three was the most accurate yet simple. Given our numerically-derived value for this slope, for each option as defined by its moneyness and maturity, we then applied (5.3) with the BS delta and vega calculated using the standard BS formulae. For the Hull and White (2017) delta we did not mimic their 36-month in-sample calibration period, which they used their empirical study of equity index options. Bitcoin options do not even have 36 months of useful data available. Besides, bitcoin prices are very much more volatile than S&P 500 index values, which another reason why we want to consider rebalancing the hedge more than once per day. Taking all these considerations into account, we used a rolling window of 30 observations at the daily frequency and 90 observations at the eight-hourly frequency to calibrate the HW delta parameters. Our results will compare the hedging error using the fixed-maturity futures, and using the perpetuals, with the HW regressions being performed twice, depending on the hedging instrument.

We shall display our results using a standard F-test for the difference in variances, using the BS delta as the benchmark, i.e. the Sticky Strike delta in (5.3). First, Table 5.1 presents the results for hedging 10-, 20- and 30-day options with moneyness between 0.7 and 1.3, and where each option is hedged with the corresponding fixed-maturity futures and rebalancing is performed every eight hours. The entries in this table and the following tables are the variance ratios, i.e. the variance of the δ^{adj}-hedging error relative to the variance of the BS delta-hedging error. The greater the effectiveness of the hedge the lower the variance of the hedging error and the efficiency gain from using the smile-adjusted delta is one minus this variance ratio. For instance, the SM/SI delta yields a variance ratio of 0.562 for hedging 10-day puts with moneyness 0.8. This indicates an efficiency gain of 43.8% compared with the BS delta-hedge, which is very significant, so the entry is marked $+++$. In the tables of variance ratios the superscripts denote the significance of one-sided F-tests on the variance ratios at 10%, 5% and 1%. For instance $***$ denotes that the δ^{adj}-hedging error has a very significantly larger variance than the BS hedging error, at the 1% level. And $++$ denotes that the δ^{adj}-hedging error has a significantly lower variance than the BS error, at the 5% level.

First consider the results for 2020 in Table 5.1. This part of the sample is characterised by slow but steady rising prices consistent with the Derman (1999) stable-trending regime when

we expect the SS delta (BS delta) to provide the most efficient delta-hedge, or the range-bounded regime when the SM delta prevails. Overall, the 2020 results in Table 5.1 display a pattern where the success of a particular delta to out-perform the BS hedge depends on the options moneyness, but not its maturity. For instance, for ATM options the ST delta is best.[11] The efficiency gain ranges from 9.7% for 30-day ATM options to 12.3% for 20-day options and 11% for 10-day options. The relative performance of the smile-implied (i.e. SM) delta is in the opposite direction to both the ST and MV deltas, not just for ATM options but for options of all moneyness. It out-performs the BS delta for OTM puts but not for OTM calls (except for 10-day calls with moneyness 1.2). For hedging 20-day deep OTM puts the efficiency gain from using the smile-implied (SM) delta throughout the whole of 2020 was $1 - 0.693 = 30.7\%$, which was very highly significant. And it was almost as high for 30-day deep OTM puts, where the efficiency gain was 28.7%. The efficiency gains from using the smile-implied hedge for other put options were much smaller, between only 3.1% and 7.6%.

Otherwise, for every other option, all the smile-adjusted deltas perform worse than the BS delta. However, this is not really surprising since the bitcoin price was stable-trending during a large part of 2020. The practical HW hedge ratio introduced by Hull and White (2017) and the minimum variance hedge of Lee (2001) also result in no improvement on the BS delta (except for ATM options). The HW delta also has a major drawback in that it uses regression to estimate its parameters which makes i.i.d. assumptions not appropriate for bitcoin which is very prone to jumps in returns. The impact of any jump remains within a rolling window for a long time and therefore has a large influence on the HW hedge ratio.

Figure 5.1 showed that 2021 was characterised by much higher and more turbulent prices, and an increase in the general level of volatility accompanied by a flatter but still asymmetric smile-shaped implied volatility curve. During the entire year of 2021 the market was characterised by huge swings in the bitcoin price as it ranged between $30,000 and almost $70,000, and as one can see from Figure 5.2 the smile at 30-days became relatively flat towards the end this period. But a flat smile makes the key ingredient the adjusted deltas, i.e. the slope of the smile, almost redundant. So it is not surprising that all smile-adjusted deltas presented no significant improvements on the standard BS hedge ratio during the second year of our sample, for all 20-day and 30-day options. However, the very short-term, 10-day smile exhibited some strange features in 2021, becoming *upward sloping* during the bull market phases of the bitcoin price. This is why the SI delta-hedge showed some very significant efficiency gains of 15.9% compared with using the BS delta for hedging 10-day OTM call options.

Next, Tables 5.2 and 5.3 examine the robustness of the results in Table 5.1 in two ways: first by repeating the analysis with rebalancing at the daily frequency (Table 5.2) and then by using the perpetual contract instead of the same-maturity futures as the hedging instrument. The results in Table 5.2 display a similar pattern to those in Table 5.1 except that they are less significant overall – but we are not surprised by this because there are now only 365

---

[11]The ST and MV delta are the same for ATM options, so the results are identical but only in this case.

instead of 1095 observations per year. They confirm our conclusion from Table 5.1 that no smile-adjusted delta can improve on the BS delta during 2021. In 2020 we also see the same pattern of performance relative to the BS delta, in that the ST delta does out-perform for ATM options but now there is some evidence that the HW deltas also beat BS, for ATM options and OTM puts with moneyness 0.9 – but none of these variance ratio statistics are statistically significant. Table 5.3 repeats exactly the same analysis as for Table 5.1, with the eight-hour rebalancing frequency, but it uses the perpetual contract as the hedging instrument for all options. We see exactly the same pattern of under- or out-performance of the BS delta as in Table 5.1, with very highly significant efficiency gains for hedging OTM puts using the SI/SM delta and the ST/MV delta for ATM options. Apart from the smile-implied delta-hedge which again provides large and significant efficiency gains for hedging 10-day OTM calls, no smile-adjusted delta significantly out-performs the BS delta in 2021. There are also some small (< 5%) efficiency gains from using the ST/MV delta for ATM options and the variance ratios are almost always smaller in Table 5.3 than they are in Table 5.1.

This finding leads us to question weather the perpetual contract provides a better hedging instrument than the futures of the same maturity as the option. To answer this question we examine variance ratios where the numerator is the variance of the perpetual-hedging error, and the denominator is the variance of the futures-hedging error. Again we divide the sample into two one-year periods, and present results by delta (now including the BS delta) and by option and Table 5.4 displays the results. In the table a variance ratio less than (greater than) one indicates that a superior (inferior) hedge is obtained using the perpetual contract. The significance of the F-statistics are marked depending on whether the perpetual provides a superior (+) or inferior (*) hedging instrument, compared with the same-maturity futures. It is clear that the results depend little on the moneyness of the option, but more on its maturity and the prevailing market state. For the 10-day options the ratios are mostly less than one for OTM calls. For 20-day and 30-day options some highly significant improvements from hedging with the perpetual are evident, especially during 2021.

Although the tables of results have provided a big picture about the overall relative efficiency of different smile-adjusted deltas, our two-year sample spans a variety of market regimes. As already observed from Figure 5.1, there are periods when the bitcoin market fluctuated quite quickly between stable-trending, range-bound and crash-jump regimes. So to help understand which delta performed best in which market state, Figure 5.4 depicts time series of the variance ratio, i.e. the variance of the smile-adjusted-delta-hedging error divided by the variance of the BS-delta hedging error. This is for rebalancing the hedge every eight hours and now each variance is calculated using only the last 90 observations – the same window as used for the HW delta parameter estimation. We emphasise that a value *greater* than one indicates that the smile-adjusted delta yields an *inferior* delta-hedging performance relative to the BS delta and we present the results on a log scale for clarity, so a variance ratio of one translates to zero in these plots. Any line below zero indicates that the delta improves

on the BS delta, but a line above zero shows the delta provides a less effective hedge than BS.

Results for 10-day options are exhibited in the upper set (a) of three plots, and results for 30-day options are exhibited in the lower set (b) of three plots. In each case (a) and (b), the top graph is for OTM put options, and these confirm the results from Table 5.1: both ST (blue) and MV (green) deltas under-perform BS for almost the entire period; as expected from the sticky classification of regimes, the SM delta out-performs the BS delta during periods when the market is range-bounded but not when it is trending, e.g. during the first big bull run starting in January 2021 and the second bull run later that year; and the performance of the HW delta is mixed. The middle graph of each set presents the variance ratios for hedging ATM options. Here all the smile-adjusted deltas are very similar because the bitcoin smile is often quite flat at this point. The bottom graph of each set shows the performance of different deltas for hedging OTM call options. Again the SM delta appears best but only for 10-day options and the improvement over BS is less than it is for OTM put options. For 30-day options no delta provides a sustained improvement over BS, especially during 2021.

Table 5.1: F-Test Hedging Results (8 Hour Rebalancing, Fixed-Maturity Futures)

Variance ratios and significance levels for one-sided F-tests on the null hypothesis $H_0 : \sigma^2_{\mathrm{BS}} = \sigma^2_{\mathrm{delta}}$ with alternative $H^+ : \sigma^2_{\mathrm{BS}} > \sigma^2_{\mathrm{delta}}$ and $H^* : \sigma^2_{\mathrm{BS}} < \sigma^2_{\mathrm{delta}}$, respectively. Hedges are based on the futures contract with the same maturity as the option, and are rebalanced every eight hours. We compare the variance of the hedging error for different deltas, relative to the variance obtained using the BS delta-hedge, and divide the two-year sample into two. We use options with three different maturities and moneyness ranging from 0.7 to 1.3, using OTM put options for moneyness < 1 and OTM calls for > 1. The 10%, 5% and 1% significance levels are denoted *, ** and *** respectively for H^*, and similarly for H^+.

| | Moneyness | | | | | | |
	0.7	0.8	0.9	1	1.1	1.2	1.3
10 Days							
2020							
ST		1.164^{***}	1.194^{***}	0.890^{++}	1.131^{**}	1.298^{***}	
SM		0.562^{+++}	0.714^{+++}	1.043	1.089^{*}	0.896^{++}	
MV		1.016	1.139^{**}	0.890^{++}	1.163^{***}	1.463^{***}	
HW		1.022	1.098^{*}	1.038	1.296^{***}	1.090^{*}	
2021							
ST		2.045^{***}	1.125^{**}	0.984	1.054	1.220^{***}	
SM		0.986	1.008	1.037	0.841^{+++}	0.841^{+++}	
MV		1.698^{***}	1.083^{*}	0.984	1.063	1.284^{***}	
HW		1.991^{***}	1.651^{***}	1.285^{***}	1.412^{***}	2.782^{***}	
20 Days							
2020							
ST	2.241^{***}	1.679^{***}	1.234^{***}	0.877^{++}	1.044	1.228^{***}	1.207^{***}
SM	0.693^{+++}	0.969	0.924^{+}	1.208^{***}	1.217^{***}	1.184^{***}	1.132^{***}
MV	1.806^{***}	1.476^{***}	1.196^{***}	0.877^{++}	1.061	1.316^{***}	1.322^{***}
HW	1.073	1.105^{**}	1.299^{***}	1.101^{*}	1.074	1.121^{**}	1.584^{***}
2021							
ST	1.972^{***}	1.630^{***}	1.067	1.046	1.037	1.100^{*}	1.179^{***}
SM	1.014	1.029	1.031	1.079	1.039	1.012	0.983
MV	1.673^{***}	1.493^{***}	1.088^{*}	1.046	1.045	1.132^{**}	1.262^{***}
HW	1.037	1.076	1.120^{**}	1.021	1.131^{**}	1.303^{***}	1.104^{**}
30 Days							
2020							
ST	2.462^{***}	1.657^{***}	1.258^{***}	0.903^{++}	1.029	1.176^{***}	1.264^{***}
SM	0.713^{+++}	0.969	0.942	1.181^{***}	1.197^{***}	1.179^{***}	1.144^{***}
MV	1.981^{***}	1.471^{***}	1.223^{***}	0.903^{++}	1.043	1.247^{***}	1.407^{***}
HW	1.001	1.299^{***}	1.454^{***}	1.172^{***}	1.342^{***}	1.361^{***}	1.369^{***}
2021							
ST	1.686^{***}	1.372^{***}	0.966	1.022	1.044	1.102^{*}	1.122^{**}
SM	1.074	1.098^{*}	1.077	1.152^{***}	1.031	0.993	0.998
MV	1.459^{***}	1.287^{***}	1.017	1.022	1.052	1.134^{**}	1.180^{***}
HW	1.024	1.194^{***}	1.241^{***}	1.202^{***}	1.264^{***}	1.135^{**}	1.134^{***}

Table 5.2: F-Test Hedging Results (Daily Rebalancing, Fixed-Maturity Futures)

Variance ratios and significance levels for one-sided F-tests on the null hypothesis $H_0 : \sigma^2_{\mathrm{BS}} = \sigma^2_{\mathrm{delta}}$ with alternative $H^+ : \sigma^2_{\mathrm{BS}} > \sigma^2_{\mathrm{delta}}$ and $H^* : \sigma^2_{\mathrm{BS}} < \sigma^2_{\mathrm{delta}}$, respectively. Hedges are based on the corresponding fixed-maturity futures and rebalanced every day. We compare the variance of the hedging error for different deltas, relative to the variance obtained using the BS delta-hedge, and divide the two-year sample into two. We use options with three different maturities and moneyness ranging from 0.7 to 1.3, using OTM put options for moneyness < 1 and OTM calls for > 1. The 10%, 5% and 1% significance levels are denoted *, ** and *** respectively for H^*, and similarly for H^+.

	Moneyness						
	0.7	0.8	0.9	1	1.1	1.2	1.3
10 Days							
2020							
ST		1.012	1.020	0.953	0.978	1.196**	
SM		0.978	1.108	1.083	1.108	0.990	
MV		1.001	1.011	0.953	0.990	1.281***	
HW		1.091	0.940	0.967	1.150*	1.159*	
2021							
ST		2.067***	1.218**	0.978	1.046	1.165*	
SM		0.987	1.085	1.087	0.992	0.975	
MV		1.737***	1.174*	0.978	1.055	1.213**	
HW		2.917***	2.324***	2.166***	2.464***	2.591***	
20 Days							
2020							
ST	2.819***	1.029	1.010	0.910+	0.999	1.206**	1.443***
SM	0.693+++	1.137	1.102	1.133	1.200**	0.983	0.840++
MV	2.213***	1.005	1.003	0.910+	1.008	1.285***	1.630***
HW	1.238**	1.134	0.958	0.992	1.109	1.401***	1.102
2021							
ST	2.013***	1.551*	1.049	0.997	1.025	1.090	1.131
SM	1.130	1.210**	1.203**	1.117	1.057	1.013	1.013
MV	1.633***	1.368***	1.031	0.997	1.032	1.120	1.195**
HW	3.271***	2.386***	2.058***	2.135***	2.700***	2.168***	1.796***
30 Days							
2020							
ST	2.699***	1.061	1.011	0.930	1.019	1.119	1.272***
SM	0.838+++	1.154*	1.090	1.117	1.142	1.179	0.930
MV	2.115***	1.026	0.959	0.930	1.028	1.172+	1.401***
HW	1.241**	1.541**	1.129	1.139	1.031	1.129	1.429
2021							
ST	1.850***	1.428***	1.022	1.015	1.027	1.093	1.113
SM	1.104	1.198**	1.177*	1.196**	1.059	1.001	1.034
MV	1.527*	1.282***	1.009	1.015	1.034	1.122	1.170
HW	3.554***	3.335***	2.534***	2.072***	2.423***	2.410***	1.852***

Table 5.3: F-Test Hedging Results (8 Hour Rebalancing, Perpetual)

Variance ratios and significance levels for one-sided F-tests on the null hypothesis $H_0 : \sigma^2_{\text{BS}} = \sigma^2_{\text{delta}}$ with alternative $H^+ : \sigma^2_{\text{BS}} > \sigma^2_{\text{delta}}$ and $H^* : \sigma^2_{\text{BS}} < \sigma^2_{\text{delta}}$, respectively. Hedges are based on the perpetual futures and rebalanced every wight hours. We compare the variance of the hedging error for different deltas, relative to the variance obtained using the BS delta-hedge, and divide the two-year sample into two. We use options with three different maturities and moneyness ranging from 0.7 to 1.3, using OTM put options for moneyness < 1 and OTM calls for > 1. The 10%, 5% and 1% significance levels are denoted *, ** and *** respectively for H^*, and similarly for H^+.

	Moneyness						
	0.7	0.8	0.9	1	1.1	1.2	1.3
10 Days							
2020							
ST		1.134^{**}	1.164^{***}	0.908^{+}	1.118^{**}	1.363^{***}	
SM		0.655^{+++}	0.709^{+++}	1.031	1.123^{**}	0.995	
MV		0.992	1.112^{**}	0.908^{+}	1.149^{**}	1.488^{***}	
HW		1.080	0.943	0.969	1.329^{***}	1.365^{***}	
2021							
ST		1.988^{***}	1.089^{*}	0.995	1.040	1.198^{***}	
SM		0.996	1.022	1.045	0.860^{+++}	0.859^{+++}	
MV		1.654^{***}	1.051	0.995	1.048	1.257^{***}	
HW		1.791^{***}	1.503^{***}	1.212^{***}	1.358^{***}	2.736^{***}	
20 Days							
2020							
ST	2.275^{***}	1.673^{***}	1.198^{***}	0.932	1.029	1.189^{***}	1.184^{***}
SM	0.742^{+++}	0.995	0.933	1.150^{**}	1.233^{***}	1.215^{***}	1.144^{**}
MV	1.846^{***}	1.478^{***}	1.166^{***}	0.932	1.045	1.267^{***}	1.290^{***}
HW	1.099	1.367^{***}	1.689^{***}	1.085	1.371^{***}	1.556^{***}	1.384^{***}
2021							
ST	1.809^{***}	1.475^{***}	0.957	0.995	1.019	1.077	1.122^{**}
SM	1.104^{*}	1.125^{**}	1.106^{**}	1.057	1.068	1.050	1.038
MV	1.548^{***}	1.370^{***}	0.985	0.995	1.025	1.105^{**}	1.184^{***}
HW	1.280^{***}	1.252^{***}	1.271^{***}	1.245^{***}	1.319^{***}	1.213^{***}	1.400^{***}
30 Days							
2020							
ST	2.364^{***}	1.597^{***}	1.213^{***}	0.941	0.997	1.138^{**}	1.231^{***}
SM	0.734^{+++}	0.986^{***}	0.965	1.174^{***}	1.214^{***}	1.195^{***}	1.151^{**}
MV	1.915^{***}	1.426^{***}	1.183^{***}	0.941	1.007	1.198^{***}	1.359^{***}
HW	1.067	1.127^{***}	1.330^{***}	1.191^{***}	1.273^{***}	1.356^{***}	1.428^{***}
2021							
ST	1.690^{***}	1.349^{***}	0.958	1.011	1.008	1.066	1.079
SM	1.190^{***}	1.241^{***}	1.206^{***}	1.160^{***}	1.075	1.036	1.057
MV	1.416^{***}	1.221^{***}	0.953	1.011	1.012	1.091^{*}	1.124^{**}
HW	1.220^{***}	1.342^{***}	1.415^{***}	1.259^{***}	1.410^{***}	1.216^{***}	1.390^{***}

Table 5.4: F-Test Futures and Perpetual Comparison (8 Hour Rebalancing)

Variance ratios and significance levels for one-sided F-tests on the null hypothesis $H_0 : \sigma^2_{\text{fut}} = \sigma^2_{\text{perp}}$ with alternative $H^+ : \sigma^2_{\text{fut}} > \sigma^2_{\text{perp}}$ and $H^* : \sigma^2_{\text{fut}} < \sigma^2_{\text{perp}}$, respectively. Hedges are based on the corresponding fixed-maturity futures and rebalanced every day. For each delta we compare two variances of the hedging error, the numerator is the variance when the perpetual is the hedging instrument and the denominator is the variance when the hedging instrument is the futures with maturity matching the option. Again we divide the two-year sample into two. We use options with three different maturities and moneyness ranging from 0.7 to 1.3, using OTM put options for moneyness < 1 and OTM calls for > 1. The 10%, 5% and 1% significance levels are denoted *, ** and *** respectively for H^*, and similarly for H^+.

| | Moneyness | | | | | | |
	0.7	0.8	0.9	1	1.1	1.2	1.3
10 Days							
2020							
BSM		1.027	1.031	1.015	0.962	0.924^{+}	
ST		1.001	1.005	1.036	0.952	0.970	
SM		1.015	1.023	1.003	0.993	0.996	
MV		1.003	1.006	1.036	0.951	0.940	
HW		1.086	0.886^{++}	0.948	0.987	1.158^{***}	
2021							
BSM		0.982	0.993	1.007	0.992	1.016	
ST		0.955	0.961	1.017	0.979	0.998	
SM		1.003	1.007	1.015	1.014	1.038	
MV		0.957	0.964	1.017	0.978	0.994	
HW		0.884^{++}	0.904^{++}	0.969	0.954	1.009	
20 Days							
2020							
BSM	0.949	0.987	1.001	0.907^{+}	0.936	0.966	0.989
ST	0.963	0.984	0.973	0.964	0.923^{+}	0.935	0.970
SM	1.016	1.014	1.011	0.864^{+++}	0.948	0.991	1.000
MV	0.970	0.989	0.976	0.964	0.922^{+}	0.930	0.965
HW	0.883^{+++}	1.132^{***}	1.225^{***}	0.951^{***}	1.108^{**}	1.254^{***}	0.902^{++}
2021							
BSM	0.896^{++}	0.870^{++}	0.864^{+++}	0.878^{++}	0.849^{+++}	0.898^{++}	0.927
ST	0.822^{+++}	0.788^{+++}	0.775^{+++}	0.835^{+++}	0.834^{+++}	0.879^{++}	0.882^{++}
SM	0.976	0.952	0.927	0.861^{+++}	0.873^{++}	0.931	0.979
MV	0.829^{+++}	0.799^{+++}	0.782^{+++}	0.835^{+++}	0.833^{+++}	0.876^{++}	0.870^{++}
HW	1.106^{*}	1.012	0.980	1.071	0.991	0.836^{+++}	1.176^{***}
30 Days							
2020							
BSM	1.003	1.000	1.022	1.057	1.025	1.022	1.026
ST	0.963	0.964	0.985	1.101	0.993	0.988	0.999
SM	1.033	1.017	1.047	1.050	1.039	1.036	1.033
MV	0.969	0.970	0.989	1.101	0.989	0.982	0.991
HW	1.070	0.867^{+++}	0.934	1.074	0.972	1.018	1.070
2021							
BSM	0.839^{+++}	0.818^{+++}	0.850^{+++}	0.910^{++}	0.867^{+++}	0.917^{+}	0.945
ST	0.841^{+++}	0.804^{+++}	0.843^{+++}	0.900^{++}	0.837^{+++}	0.887^{++}	0.909^{+}
SM	0.930	0.924^{+}	0.952	0.917^{+}	0.905^{+++}	0.956	0.991
MV	0.815^{+++}	0.776^{+++}	0.797^{+++}	0.900^{++}	0.835^{+++}	0.882^{++}	0.901^{++}
HW	0.999	0.920^{+}	0.969	0.953	0.968	0.983	1.158^{***}

Figure 5.4: Hedging Performance on a Rolling Sample

Variance ratios showing performance of the various perpetuals hedge ratios relative to the BS delta, using 8-hourly rebalancing, where variances of the hedging errors are calculated using the previous 90 observations. We present results for (a) 10-day and (b) 30-day options over the two-year sample. The solid line at 0 acts as a reference value, with ratios above 0 indicating inferior performance relative to BS and ratios below 0 indicating superior performance relative to BS. The top graph in (a) depicts the performance for OTM put options with $m = 0.8$ and for (b) put options with $m = 0.7$, the middle graph shows ATM options for both (a) and (b), and the bottom one is OTM call options with (a) moneyness 1.2 and (b) moneyness 1.3.

(a) Results for 10-day Options
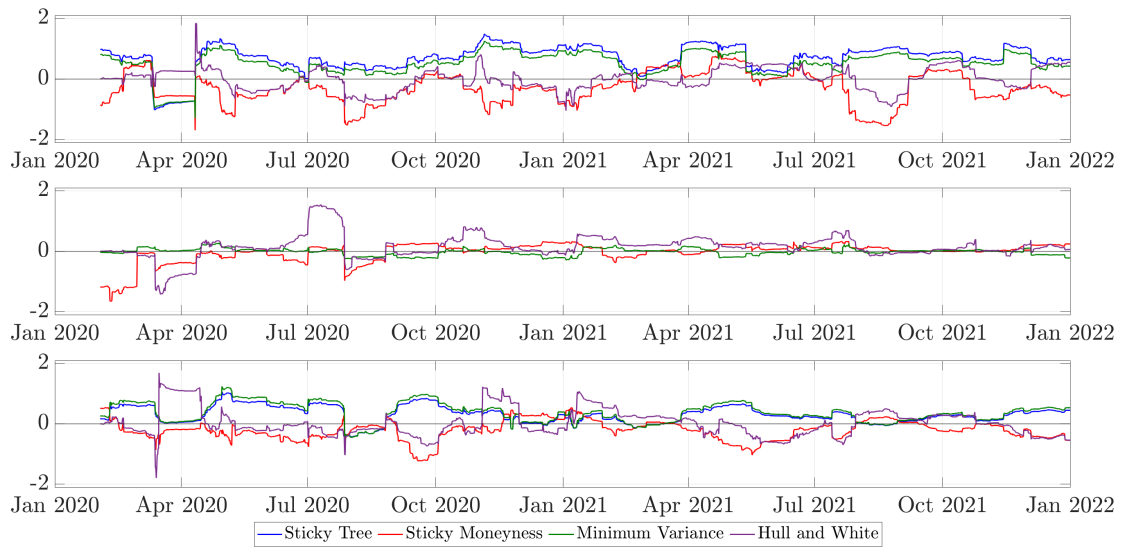
(b) Results for 30-day Options
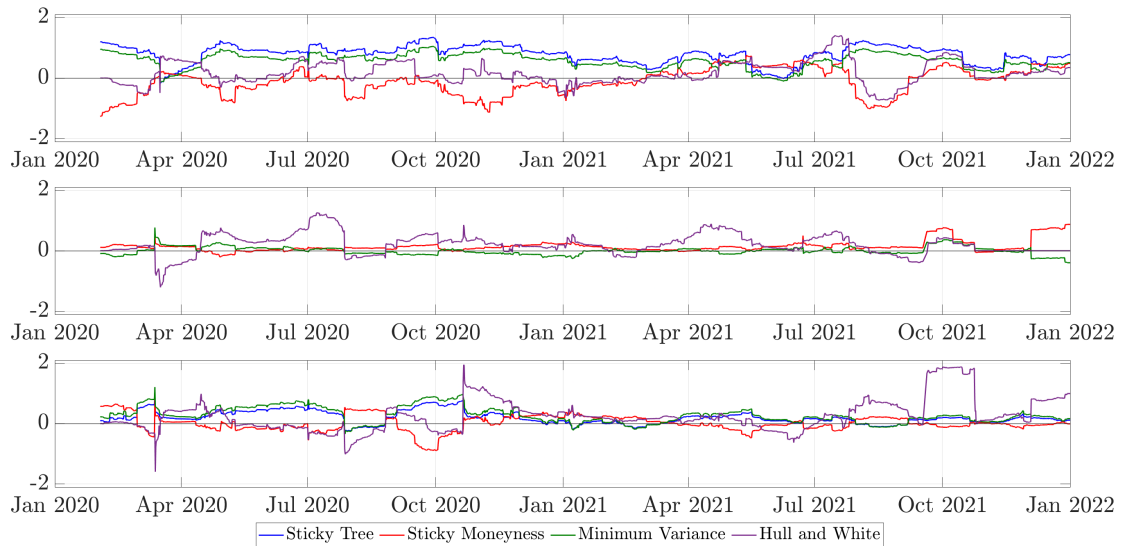
5.5 Concluding Remarks

We compare the hedging effectiveness for the most actively traded bitcoin options on the Deribit exchange, i.e. options with strike levels ranging 30% above and below the current BTC index and with expiry up to one month. We analyse the variance of the delta-hedging errors where the hedging instrument is either the futures of the same maturity as the option, or the perpetual contract – an innovative product that is unique to cryptocurrency derivatives markets. With rebalancing of the hedge either every eight hours (to coincide with funding payments on the perpetual) or daily, and using either the same-maturity futures or the perpetual as the hedging instrument, we find some very robust results. Also, rather than a simple tabular comparison of the mean square errors from different hedge ratios, we have applied a simple variance-ratio test which provides the statistical significance of efficiency gain from using a given delta, relative to the BS delta.

This way we have demonstrated that the smile-implied (sticky-moneyness) delta can provide a significantly better hedge than a standard Black-Scholes delta for out-of-the-money options, with efficiency gains of over 40% in some cases. The minimum-variance delta is also better than the BS delta, but only for at-the-money options, where it coincides with the sticky-tree delta. No other smile-adjusted delta can improve on the Black-Scholes delta consistently, and even the smile-implied and minimum-variance delta-hedge performance were poor during much of 2021. The exception is the smile-implied hedge for short-term out-of-the-money calls, at times when the slope of the implied volatility curve became positive. In contrast to equity indices like the S&P 500, the bitcoin price does not trends upwards in a stable fashion and then suddenly crash – its upwards price jumps can be as large as its downward price jumps so its smile can be quite symmetric – or even completely upwards sloping. We have also shown that the perpetual contract is a significantly better hedging instrument than the same-maturity futures as the option, irrespective of the option's moneyness. This is particularly evident for options with longer maturity, where the basis between the perpetual and the futures is greatest.

The focus of this chapter on dynamic delta-hedging with frequent rebalancing may help market makers in bitcoin options gain a competitive edge in a market that only started maturing in 2021. However, at the time of writing the trading costs of hedging price and volatility risk with options might erode any extra profits made from potentially increasing trading volumes through a reduction in spreads.

CHAPTER 6

IMPLIED VOLATILITY INDEX

Motivation: One of the primary attributes of bitcoin, and a principal motivation to trade it, is indisputably its pronounced volatility. Nevertheless, until recently, the direct trading of bitcoin volatility was solely achievable through a combination of options and spot. Standard tools for volatility trading of traditional markets, such as variance swaps or futures on volatility indices were yet to be introduced. A bitcoin volatility index offers invaluable insight into the prevailing options market. Such an index summarises the future volatility expectations of sophisticated option traders. It gauges market sentiment, enables traders to mitigate and manage risk and provides an underlying to trade volatility directly. More important, the index allows to examine empirical properties of the bitcoin variance risk premium, thereby fostering a comprehensive understanding of the bitcoin options market.

Summary: We present a bitcoin implied volatility index, the first of its kind, using a variance swap fair-value formula. Motivated by liquidity and trading clusters, we construct bitcoin implied volatility indices with different maturities at daily frequency and investigate the BTC realised volatility. Subsequently, we discuss the features of the index and the associated bitcoin variance risk premium, and examine the relationship between bitcoin's monthly realised variance, the volatility index, and the variance risk premium, compared to their equivalents for US equities, oil, gold, and the EURUSD exchange rate. Surprisingly, we find bitcoin VRP behaves like other asset classes and showed over long periods the most attractive VRP profile for swap writers.

6.1 Trading Volatility

There are three distinct approaches to trade volatility of traditional assets, which are also partially applicable to cryptocurrency markets: (i) trading delta-hedged options or specific option strategies; (ii) entering volatility or variance swap agreements; and (iii), using futures on implied volatility indices. The delta-hedging strategies discussed in Chapter 5 present the most rudimentary form of volatility trading, as it eliminates the underlying bitcoin exposure completely. Assuming all input factors, aside from the underlying, remain constant, the payoff of delta-hedged options depends on the variance of the underlying. If an investor believes the implied volatility is lower (higher) than the realised volatility throughout the option's lifespan, he can profit by buying (selling) volatility, continuously re-hedge the position and lock in profits. This technique, often referred to as gamma scalping, capitalises on the option's carry, i.e. the difference between implied and realised volatility. Other option strategies, such as straddles, strangles or butterflies allow to take directional positions in future realised volatility. Note what all these strategies have in common is that they are generally delta-neutral, or delta flat, at inception, but are influenced by – among others – gamma, vega and theta.

In theory, having knowledge of future volatility and the capacity to continuously hedge would result in a fixed profit for a delta-hedged option, which would be based solely on the carry. However, in practice, the combination of discrete delta-hedging and unknown future volatility prevents this simplified assumption and led to the introduction of swap derivatives. Particularly, volatility, variance or gamma swaps are prominent and liquid derivatives in equity markets, providing pure volatility exposure and remove the need to continuously delta-hedge an option.[1] The payoff of volatility and variance swaps at maturity T are straight forward:

$$(RV - \sigma_{SR}) \times \text{notional} \qquad \qquad \text{(Volatility Swap)}$$

$$\left(RV^2 - \sigma_{SR}^2\right) \times \text{notional} \qquad \qquad \text{(Variance Swap)},$$

where RV denotes the Realised Volatility (RV) over the life of the contract and σ_{SR} reflects the swap rate volatility which is fixed at the start of the contract. Assuming the underlying follows a geometric diffusion process, similar to (4.3.2), the continuously monitored realised variance RV^2 from time t to T is given by:

$$RV^2 = (T - t)^{-1}\Sigma_{t,T} \quad \text{where} \quad \Sigma_{t,T} = \int_t^T \sigma_u^2 du,$$

and thus the risk-neutral fair value swap rate is given as:

$$\sigma_{SR}^2 = (T - t)^{-1}\,\mathbb{E}\left[\Sigma_{t,T}\right].$$

Hence, these derivatives allow investors to trade future realised volatility (variance) against

---

[1]The term *swaps* is rather misleading as these derivatives have fixed maturities and no payments in-between. Hence, they resemble more forward contracts.

current implied volatility (variance) with the same maturity. As such, they are designed to have a constant exposure to volatility, resulting in a flat vega profile that appeals to traders. While these swaps enjoy great popularity for equity, their trading for bitcoin remains nascent. At the time of writing, variance swaps for cryptocurrencies are traded OTC, or on-chain only, if at all. Indeed, only the digital asset trading firm GSR officially launched the first on-chain variance swap but expressed concerns regarding insufficient interest and liquidity.[2] As discussed in the literature review, a variance swap can be replicated through a static options portfolio with same maturity, wherein each option is weighted by the squared inverse strike (Neuberger, 1994; Demeterfi et al., 1999), and does not require further transactions throughout the span of the swap.[3] As such, these weightings make the options portfolio take a long position in implied skew, as there is more exposure to downside put than to upside calls; as well as convexity, due to the greater weight on the wings. However, the limitations of this replication are self-evident: The assumption that all strikes can be bought is not assured. Particularly OTM options – and even stronger for cryptocurrency options – show low liquidity which leave any writer of these swaps vulnerable to short tail risk.

The variance swaps replication serves as the fundamental methodology of modern volatility indices, which complete our introduction of most common volatility trading tools. It provides information as a risk-neutral expectation of the S&P500 variance and its interpretation of both stock market uncertainty, i.e. the 'physical' expected volatility, and as variance risk premium, which is the expected premium from selling variance swap contracts. A volatility index summarises, in a single value, the information embedded within the implied volatility smile for a specified maturity. [4] However, there exist substantial disparities between volatility indices and variance swaps. First, volatility indices commonly exclude extremely high and low strikes, thereby enhancing stability but undermining the significance of deep OTM puts. Second, variance swaps are usually priced using a continuous implied volatility surface, whereas the discrete sampling of listed options and the noise due to rolling expiries introduces small discrepancies. Finally, a linear interpolation between expiries assumes a flat volatility term structure which is not evident in practice, neither in equity nor cryptocurrency markets, see Bennett (2014) for an in-depth analysis. The variance risk premium may be defined as the reward required by a risk-averse investor to compensate for the risk of both stochastic volatility and jumps in price of the underlying asset, see Todorov (2010) and Konstantinidi and Skiadopoulos (2016a).

Typically, soon after an exchange quotes a volatility index they begin listing futures and options on it. These exchange-traded derivatives allow large institutions to trade volatility as an asset class in its own right instead of indirect trades such as option straddles which

---

[2]See GSR Variance Swaps for more details.

[3]However, this formula is only an approximation of the realised variance as the integral over the strike range, discussed in Chapter 2, goes from zero to infinity which is practically not possible.

[4]Theoretically, the index signifies a fair value for a variance swap that exchanges a fixed for a floating realised variance over the life of the swap. Consequently, these indices measure the variance risk premium across different investment horizons.

require constant and costly rebalancing.[5] This development is currently observable within the cryptocurrency market. Our work in this chapter marked the pioneering introduction and highlighted the importance and necessity of an implied volatility index for bitcoin. Following the publication of our study, Deribit launched their own Deribit Implied Volatility Index (DVOL), which is based on and expands our methodology.[6] A year later, in March 2023, the exchange initiated futures trading on this novel underlying, allowing traders to directly trade bitcoin volatility directly without resorting to costly option transactions. As such, Deribit is the sole exchange offering listed futures on any cryptocurrency implied volatility index. Although the trading volumes and open interest of these futures remain very low, and their diversification or hedge potential is not as robust as for equities, as discussed in Section 5.2, we believe that offering these futures constitutes the appropriate and logical next step in the continuously maturing and evolving bitcoin derivatives market.

This chapter introduces volatility indices for bitcoin with various maturities between ten days and four months. Having explored the time series properties of our bitcoin VIX indices, we use them as fair-value variance swap rates for estimating bitcoin variance risk premia of different maturities. As elaborated in Chapter 2, a substantial body of empirical research literature analyses the variance risk premia in different asset classes. This risk premium is typically negative, as investors are willing to pay for the highly effective diversification afforded by including volatility in a portfolio – variance usually has a very high negative correlation with equity returns. Contrary to expectations, our findings suggest that the bitcoin VRP does not adhere to the anticipated pattern of low negative payoffs with infrequent yet sudden jumps with indications of option overpricing. Additionally, when compared to other asset classes, our analysis reveals that bitcoin has been the most lucrative option for any variance swap writer, and it continues to outperform the S&P 500 index as a more attractive underlying asset for writing a variance swap. Finally, our study uncovers evidence suggesting a significant relationship between the implied volatility indices of gold and FX markets, and bitcoin.

6.2 Methodology

Initially, volatility indices were calculated as the mean of the implied volatilities of eight distinct ATM options (Carr and Wu, 2004). The CBOE was the first exchange to introduce an ATM volatility index, adhering to the methodology established by Whaley (2000). For each maturity, the implied volatility of two call and put options with strike levels encompassing the underlying value are averaged, followed by a linear interpolation to derive an ATM implied volatility. Subsequently, these two ATM implied volatilities for varying maturities are

---

[5]The usual next development is to bring this volatility trading to individual speculators via exchange-traded funds and notes – albeit with extremely high risk Alexander and Korovilas (2013).

[6]See Deribit Volatility Index for more details.

interpolated once more to obtain a fixed-maturity ATM implied volatility.[7] The final index (VXO) was the equally weighted volatility between the two straddling maturities. As such, it was an estimate of the one-month ATM implied volatility and an accurate approximation of the volatility swap rate.

This methodology was later updated as soon as practitioners started to realise that variance, not volatility was the correct measure of deviation.[8] Today, most provider use a variance-based calculations, e.g. vStoxx (V2X), VSMI (V3X) or CBOE VIX index. However, a great number of researchers (Leontsinis and Alexander, 2017; Griffin and Shams, 2018) underscore the limitations of variance swap-based calculations, including jump errors and susceptibility to manipulation, and advocate for more robust methodologies. The proposed bitcoin implied volatility index follows the same variance swap-based calculation, and we shall explore its characteristics, as well as alternative index calculations, in greater depth in the next section.

All existing volatility indices pertain a constant 30 days maturity, and are frequently described as a representation of option traders' expectations of average volatility on the underlying over the next calendar month. However, it is possible to modify the methodology to quote indices of any constant maturity on any underlying asset or index, provided there is sufficient trading volume on the options. Alternatively, for assets such as bitcoin whose market prices move much more quickly than equities, indices that represent expectations of average volatility over the next days or weeks may be of interest. Hence, we do not restrict ourselves to these maturities. As discussed in Section 3.2, the majority of active trading occurs on options set to expire within the next month. As such, we examine short (10 days), mid (20 days), and long (30 days) maturities, which collectively account for 80% of the aggregated trading volume. Moreover, diverging from the hedging study in Chapter 5, we explore ultra-long maturities of 60, 90, and 120 days, despite the diminished active trading and liquidity associated with these maturities. Furthermore, as seen in Section 3.2, these maturities hold most of the open interest. An investor holding an option with an ultra-long maturity may perceive an implied volatility derivative as the optimal hedge, in lieu of engaging in costly, frequent rebalancing.

---

[7]Additionally, the CBOE implemented an artificial "trading-day conversion" that contrasts the actual 365-day annualisation of the Black-Scholes implied volatility, and converts it to 'trading-day' volatility:

$$TV(\tau) = \frac{\sigma_t(1,\tau)\sqrt{NC}}{\sqrt{NT}},$$

where $\sigma_t(1,\tau)$ is the BSM ATM IV with time to maturity τ, NC is the number of calendar days and NT number of trading days.

[8]Bennett (2014) points out three main reason why variance, not volatility should be considered as the measure of deviation: (i) Variance takes into account the implied volatility of all strike levels for option with same maturity; (ii) the sum of squared deviations, i.e. the variance, is usually taken to avoid cancelling; and (iii), the profit of a delta-hedged option depends on the squared returns of the underlying and hence variance is a better measure.

6.2.1 Variance-Swap Based Calculation

We introduce the family of bitcoin implied volatility indices, which we call the BVIN. We call indices based on the CBOE VIX methodology 'geometric' not because they assume the underlying follows a geometric Brownian motion (without jumps) but because they represent the fair-value variance swap rate for a realised variance defined as the average sum of squared log price changes over the life of the swap. We follow the exact calculation of CBOE Global Markets (2003) assuming the risk-free rate r is zero, which makes virtually no difference to our calculations, and employ a Riemann sum to approximate the fair-value integral. That is, for $j = 1, 2$ we set:[9]

$$\theta_{jt} = 2 \sum_{i=1}^{k_j} K_i^{-2} f(K_i, T_{jt} \mid F_{T_{jt},t}) \Delta K_i - \left(\frac{F_{T_{jt},t}}{K_0} - 1 \right)^2, \tag{6.1}$$

where T_{1t} is the actual traded maturity closest to but *less* than or equal to the target constant maturity T and T_{2t} is the actual traded maturity closest to but *greater* than T; k_j is the number of options available at maturity T_{jt}; $F_{T_{jt},t}$ is the price of the forward contract with maturity T_{jt} derived from the option prices; $f(K_i, T_{jt} \mid F_{T_{jt},t})$ is the price of the OTM option with the same maturity and strike K_i; K_{0j} is the separation strike, i.e. the first strike below the forward price – we use puts for strikes $K_i < K_{0j}$ and calls for strikes $K_i > K_{0j}$;[10] and $\Delta K_i = \frac{K_{i+1} - K_{i-1}}{2}$. Then if the (constant) maturity T is measured in days, the geometric index at running time t is calculated as:

$$V_t^T = \sqrt{\omega_t^T \theta_{1t} + (1 - \omega_t^T) \theta_{2t}} \times \sqrt{365/T}, \tag{6.2}$$

with

$$\omega_t^T = \frac{n_{2t} - n}{n_{2t} - n_{1t}},$$

where n is the number of minutes for the constant maturity and n_{jt} is the number of minutes until maturity T_{jt}, for $j = 1, 2$. In equation (6.1) K_{0j} is the model price of the futures inferred from put-call parity. The difference between K_0 and $F_{t,T}$ is minuscule. Therefore, the second term in (6.1) only has a tiny effect on the BVIN indices and we can ignore it.

Having constructed the BVIN indices for different maturities, we use the T-maturity index at time t as the fair-value swap rate to derive a VRP for maturity T, at time t. We use the average squared log return on the bitcoin spot over the life of the swap to compute the realised volatility (RV) and equate the geometric VRP^G to the variance swap payoff:

$$\text{VRP}_{t,T}^G = \text{N} \left[\text{RV}_{t,T}^2 - \text{BVIN}_{t,T}^2 \right], \tag{6.3}$$

---

[9]We omit the division by maturity T_{jt} here, although the implied variance formula is often presented in that (annualised) form. This is because the subsequent formulae are simplified when we annualise only at the end result, as in (6.2).

[10]To avoid overcomplicating the notation we assume the strikes are all of the same maturity.

where N is a nominal amount in USD, with realised volatility defined as:[11]

$$\text{RV}_{t,T} = \sqrt{A\,M^{-1} \sum_{i=0}^{M-1} r_{t+i}^2}, \tag{6.4}$$

where A is the annualising factor, M is the number of time-intervals in the constant maturity T and r_t is the log return from time t to $t+1$, based on a partition of $[0,T]$ into M intervals.

A comprehensive body of literature, both theoretical (Griffin and Shams, 2018) and practical (Bennett, 2014), underscores the susceptibility of such calculations to index and futures settlement manipulations. As the calculation is heavily tilted towards deep OTM put, only relatively minimal effort would be required to manipulate these few downside puts and subsequently impact the entire index. In response to this issue, other indices employing similar calculations, such as vStoxx, have transitioned to use a 30-minute average ending price. Nevertheless, this presents a considerable challenge for bitcoin options, as the number of traded contracts used to generate this type of index is still very small compared to US equity. An attacker would merely need to manipulate the price of two, or at most four, deep OTM put options to attain the desired index value. To tackle this issue, we consider an arithmetic index calculation proposed by Leontsinis and Alexander (2017), which is less prone to manipulation.

6.2.2 Arithmetic Calculation

In this sub-section, we can examine the disparities between the CBOE methodology and an index based on the 'arithmetic' variance swap rates of Leontsinis and Alexander (2017). The authors deduce a formula for the fair-value of a realised variance, defined as the average sum of squared price changes, rather than the change in log prices. Their motivation lies in employing a definition that satisfies the aggregation property introduced by Neuberger (2012) and later generalised by Alexander and Rauch (2021).[12] The benefit of defining realised variance this way is that the fair-values have no jump bias. By contrast, in the presence of jumps, the standard 'geometric' formula used by the CBOE can have very large errors (Aït-Sahalia et al., 2020). If bitcoin prices jump, the arithmetic index would be a more accurate fair value, and also be a better, unbiased, measure of the bitcoin variance risk premium. As discussed in Chapter 2, bitcoin is prone to jumps, justifying the application of the arithmetic alongside the geometric calculation. This would be particularly significant for longer-term premia, spanning the next month or beyond, as the impact of jumps accumulates with maturity.

This discussion motivates the use of the arithmetic variance swap rate introduced by Leontsinis and Alexander (2017) which has no jump error, being valid for any martingale

---

[11]Note that there are other methodologies to calculate the realised variance. However, we refer to the standard market convention using discrete squared log-returns as outline in J.P. Morgan variance swap pricing document.

[12]The aggregation property describes a characteristic of financial assets that measures the volatility (or any other characteristic) of the underlying. It states that the sum of volatilities over any non-overlapping time intervals should be equal to to volatility of the entire combined period.

underlying price process, such as bitcoin futures. The resulting index, denoted AVXB (AV_t^T) is calculated using the same interpolation (6.2) but now setting:

$$\zeta_{jt} = 2F_{T_{jt},t}^{-2} \sum_{i=1}^{k_j} f(K_i, T_{jt} \mid F_{T_{jt},t})\Delta K_i. \tag{6.5}$$

and

$$AV_t^T = \sqrt{\omega_t^T \zeta_{1t} + (1 - \omega_t^T)\zeta_{2t}} \times \sqrt{365/T}, \tag{6.6}$$

where we use the same notation as for equation (6.1) and (6.2).

By contrast with (6.1), (6.5) aligns with a realised variance definition that satisfies the aggregation property introduced by Neuberger (2012). As a result, the arithmetic fair value remains unaffected by the jump biases that present challenges for the standard calculation. The only assumption required is that the underlying asset follows a martingale process. The primary distinction between the two formulas is the use of futures prices rather than option strikes in the Riemann sum. Consequently, this formula offers the advantage of being truly model-free while eliminating distortions that impact the VIX when low-strike options (i.e., deep OTM puts) possess stale prices.[13] Additionally, the formula (6.5) is less susceptible to manipulation via large buy or sell pre-settlement orders on illiquid low-strike options, which have been previously used to manipulate the VIX (Griffin and Shams, 2018), as briefly mentioned in the preceding section. While there is no evidence of price manipulation in Deribit options, there are indications that liquidity of OTM puts diminishes when volatility rises – precisely when issuers of bitcoin variance swaps require their hedges the most. Analogous to the geometric VRP calculation, we define the arithmetic VRP^A as follows:

$$\text{VRP}_{t,T}^A = \text{N}\left[\text{ARV}_{t,T}^2 - \text{AV}_{t,T}^2\right], \tag{6.7}$$

where ARV is the realised volatility calculated using changes in the underlying price rather than log-returns as advocated in Leontsinis and Alexander (2017).

6.3   Bitcoin Volatility Index and its Risk Premia

In this section, we present an empirical analysis of the two index calculation methodologies outlined earlier, assessing their performance as fair variance swap rates for bitcoin variance swaps across various maturities. The first subsection examines the BVIN, which corresponds to the geometric variance-swap based indices with maturities of 10, 20, 30, 60, 90, and 120 days. The motivation behind investigating the short-term maturities lies in the rapid price

---

[13]The VIX and other equity index volatility indices are normally calculated using daily close option prices. That is, the price of the last recorded trade of the day. But this could be hours old, especially for deep OTM options where trading volume is often lower than for near ATM options. But the VIX formula places more weight on deep OTM puts than it does on other options, and this feature is well-known to produce distortions in the VIX, and other similar indices.

fluctuations inherent in bitcoin markets, which significantly surpass the volatility observed in fiat currencies, conventional commodities, or other established capital markets. Additionally, Deribit bitcoin options tend to be issued at shorter maturities than most other options, allowing for the calculation of BVIN indices at shorter maturities than any other implied volatility indices. The motivation for analysing longer date maturities stems from the open interest clustering observed for quarterly maturities and the potential hedging purpose of the implied volatility index. In the second subsection, we evaluate the performance and characteristics of the arithmetic volatility index and draw comparisons between the arithmetic and geometric volatility indices, and their associated risk premia.
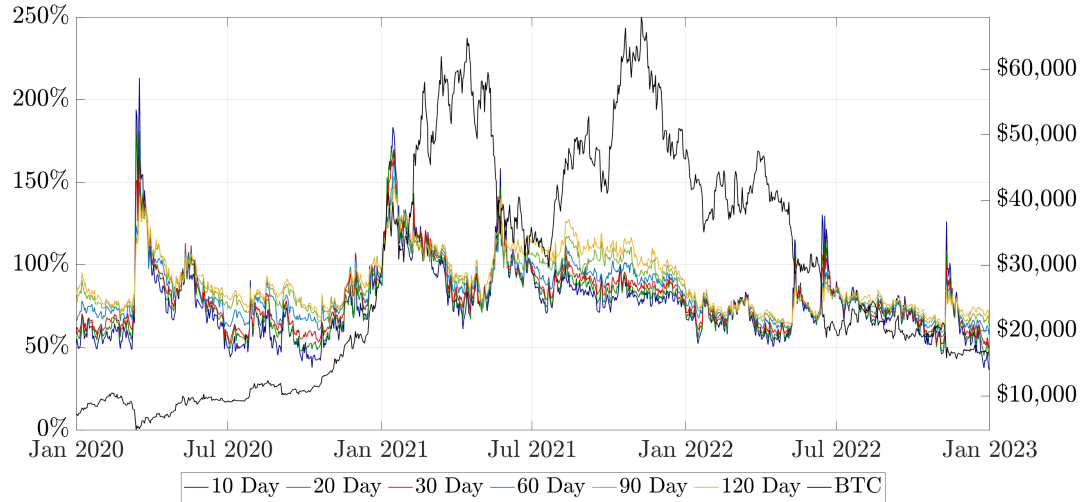
6.3.1 The BVIN

Let us consider first Figure 6.1, which illustrates the daily evolution of the bitcoin implied volatility indices with different fixed maturities, from January 2020 to 2023. Each data point is calculated using (6.1) and (6.2), using the whole chain of option prices. The term structure of the BVIN starts with contango during an initial three-month stable period up to 12 March 2020, with the 10-day BVIN at approximately 55% and the 120 day BVIN closer to 70%. On black Thursday, when the COVID-19 crisis reverberated through the financial landscape, bitcoin's value plummeted by over 30% in a single day. This systematic risk shifted the BVIN term structure into backwardation, at record levels between 130% for the 120-day and 220% for the 10-day BVIN, respectively. As the market transitioned into a more range-bounded period from June to November 2020, the term structure reverted to contango around the 60% level. This pattern recurs quite frequently; generally, the term structure remains in contango with with varying magnitudes. Following sudden price jumps, such as those in January and May 2021 (Crypto Winter), and May and July 2022, the term structure transitions to (strong) backwardation, before correcting in a mean-reverting manner back to contango. The only exception being Spring 2021, after Bitcoin's bull run, where the positive correlation between the underlying and implied volatility causes the BVIN to decline from approximately 150% to 70%, while maintaining the backwardation structure. This behavior is characteristic of many other implied volatility indices, such as the S&P 500 VIX and other VIX-type indices analyzed in Section 6.4. Towards the end of the sample period, the curve remains in a calm contango around the 60% volatility level, with differences between short- and long-term maturities reaching highest recorded levels. In fact, since July 2021, excluding the systematic impacts on the underlying (May 2022, TerraLuna Crash; June/July, missed CPI target; November 2022, FTX collapse), bitcoin's implied volatility has been steadily declining, indicating a maturing market with increasing liquidity.

Now, let us turn our attention to the realised volatility calculations based on (6.4). Figure 6.2 illustrates exemplary the 30-day BVIN (depicted in red) alongside the 30-day forward looking realised volatility (shown in black) on a rolling window.[14] Similar to the declining

---

[14]We do not depict other maturities as they all follow a rather similar structure, not adding much more

Figure 6.1: Bitcoin Implied Volatility Indices

Bitcoin implied volatility indices for different maturities constructed from out-of-the-money put and call bitcoin option prices using formula (6.1) for 10 (blue), 20 (green), 30 (red), 60 (cyan), 90 (light green) and 120 (orange) days maturity. Each time series is constructed at daily frequency between January 2020 to 2023 and we present volatilities in percentage points.

observation for the BVIN, we find the same declining tendency for the RV. In fact, since the peak of the bull-run in Spring 2021, the RV exhibits a clear downward trend. The BVIN is generally higher than the RV across our three-year sample, falling below the RV on only eight exceptional occasions which include price jumps due to systematic risk or unprecedented (frequent) change of spot-vol correlation. This suggests that bitcoin options generally have a positive carry, implying that volatility is overpriced. Indeed, it is only for the COVID crisis that the RV exceeds the BVIN by more than 140%; for the other seven occasions, we find relatively low discrepancies ranging between 3% and 30%. The duration during which the RV exceeds the BVIN varies significantly as well. It is only during the pandemic (March 2020) and the FTX (November 2022) crash that the RV surpasses the index throughout the entire 30-day span. In all other instances, it ranges between two days and three weeks. The sudden jumps in RV in March 2020 or November 2022 are statistical artefacts or 'ghost-features'.

Figure 6.3 depicts the VRP calculated using (6.3), with RV derived from daily log returns for all maturities. The forward-looking nature of RV is evident here, as the k-day VRP experiences an upward jump precisely k days prior to a substantial positive or negative return on bitcoin. The COVID crisis is frequently referred to as a 'once in a lifetime' event, and as such, the VRP for short maturities reached approximately 100,000 USD. That is, an issuer of a bitcoin variance swap with 1 USD notional would be obligated to pay out 100,000 USD on any swaps maturing at this time. Longer-term VRPs exhibit less variability and have been

_____

information.

Figure 6.2: 30-Day BVIN and Realised Volatility

The 30-day BVIN calculated using (6.1) and forward looking 30-day realised volatility using (6.4). We keep the same color for the BVIN and depict the RV in black on a daily frequency between January 2020 to 2023. All values are presented in annualised percentage points.
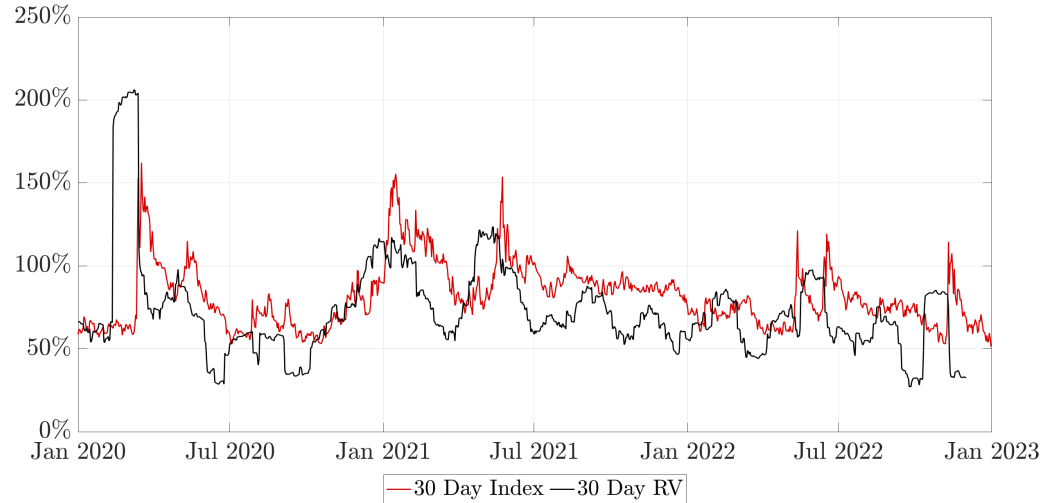
predominantly negative since January 2021, with the buyers of realised variance being required to pay the issuer anything between 100 to 8,000 USD, primarily dependent on the maturity date. In fact, short-term swaps transition into positive territory more frequently, exactly 18 times within our three-year sample. In contrast, the 30-day swap only enter positive terrain in eleven instances, and the 120-day swap merely four times, making short-term swaps less appealing for writers. A noteworthy characteristic of Figure 6.3 is the increased stability of the bitcoin VRP from July 2021 onwards, displaying a striking resemblance to an equity VRP in recent times. Consequently, the VRP remains negative most of the time, only transitioning to positive for a few days when the bitcoin VIX term structure enters into backwardation.

Figure 6.3: Bitcoin Variance Risk Premia

The bitcoin variance risk premia for for different maturities, derived using formula (6.3) with daily monitoring of the forward looking realised variance from January 2020 to January 2023. We set the notional amount $N = \$1$ and keep the same colors for different maturities. Fo the sake of clarity, we break the payoff axis (y-axis) between \$24,000 and \$48,000.
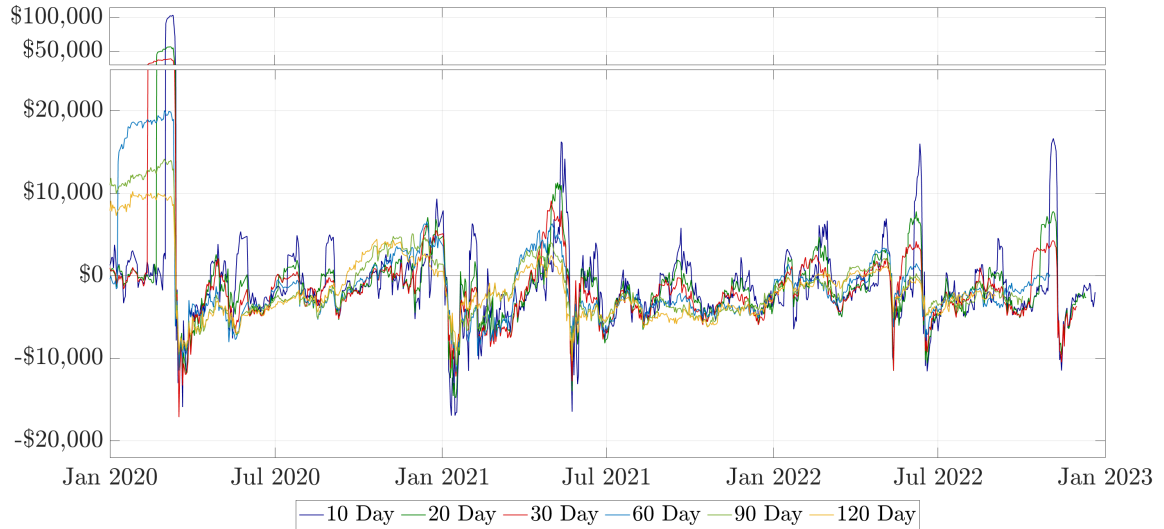
6.3.2 Arithmetic Index

In this subsection, we focus on the arithmetic index calculation by Leontsinis and Alexander (2017) for bitcoin options data. Figure 6.4 depicts the arithmetic bitcoin VIX indices (upper plot), and the difference between the BVIN and AVXT $(6.1) - (6.5)$ (lower plot), both in percentage points of volatility. The arithmetic index exhibits similar patterns and trends to the geometric calculation, with the term structure primarily in contango but transitioning to backwardation following abrupt price jumps. A notewothy difference are the distinct magnitudes in both contango and backwardation between the BVIN and AVXT. As expected, the arithmetic indices are generally higher than the geometric BVIN due to the negative jump bias in the latter. This difference becomes more prominent during particularly volatile periods, such as March 2020 and summer 2021, and increases with maturity because the likelihood of a price jump before the swap expires also increases. For instance, the difference for the 120-day swap reaches up to 22 percentage points. Interestingly, maturities up to 30 days demonstrate only minor discrepancies between the two calculation methods, rarely exceeding $\pm 1\%$, relative to the BVIN. This can be attributed to the diminished impact of the jump error for shorter maturities. Another contributing factor is that the strike range traded expands with maturity, especially for deep OTM puts. The geometric BVIN assigns greater weight to these low-strike options, which also have considerably lower prices than ATM options. Nonetheless, there appears to be a shift in future expectations as the difference between the geometric and arithmetic approaches increasingly zero. The arithmetic index follows the de-

clining volatility trend and exhibits a less pronounced backwardation. The two crypto-specific crashes (TerraLuna and FTX) result in slightly lower arithmetic indices compared to their geometric counterparts, a discrepancy that is corrected between the two crashes.

Next, we examine the arithmetic index value as the fair variance swap rate. The top half of Figure 6.5 presents the VRP for a \$1 notional using (6.7), while the bottom illustrates the difference between the geometric and arithmetic VRP (6.3 - 6.7). Initially, both calculations seem to yield similar outcomes in terms of absolute values and patterns. In fact, the difference between swap maturities up to one month is nearly negligible, excluding unanticipated market crashes. The number of jumps into positive payoff for the swaps is identical to those of the geometric VRP, with the sole exception being the 30-day swap being positive nine times instead of eleven. The primary differences arise from the large discrepancies of the AVXT. Specifically, longer-term swaps remain negative for extended periods and are generally smaller than their geometric counterparts. We conclude that, for the 30-day variance swap rate, the arithmetic calculation captures the expected future realised variance more accurately, as its deviation from the RV is smaller than the geometric calculation. Longer maturities tend to overestimate the RV, and thus are less likely to result in a trade.

Figure 6.4: Arithmetic Bitcoin Implied Volatility Indices & BVIN Differences

Arithmetic bitcoin implied volatility indices for different maturities constructed from out-of-the-money put and call bitcoin option prices using formula (6.5) for 10 (blue), 20 (green), 30 (red), 60 (cyan), 90 (light green) and 120 (orange) days maturity (upper plot); and the difference (BVIN-AV) on the lower figure. Each time series is constructed at daily frequency between January 2020 to 2023 and we present volatilities and differences in percentage points.
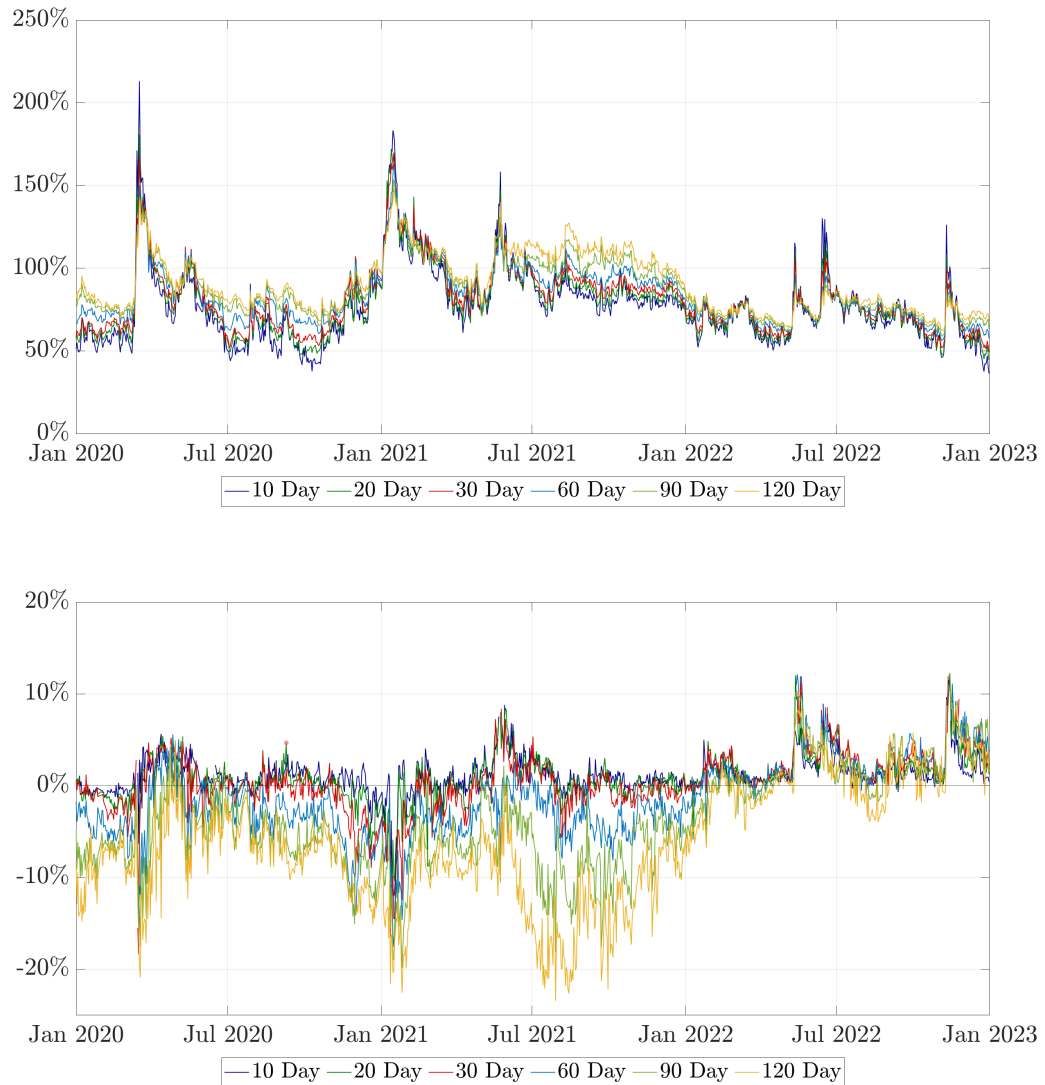
Figure 6.5: Arithmetic Bitcoin Variance Risk Premia

Arithmetic bitcoin variance risk for different swap maturities with $1 notional using (6.7) for 10 (blue), 20 (green), 30 (red), 60 (cyan), 90 (light green) and 120 (orange) days maturity (upper plot); and the difference $(\text{VRP}^G\text{-VRP}^A)$ on the lower figure. Each time series is constructed at daily frequency between January 2020 to 2023 and we the payoff and differences in USD.
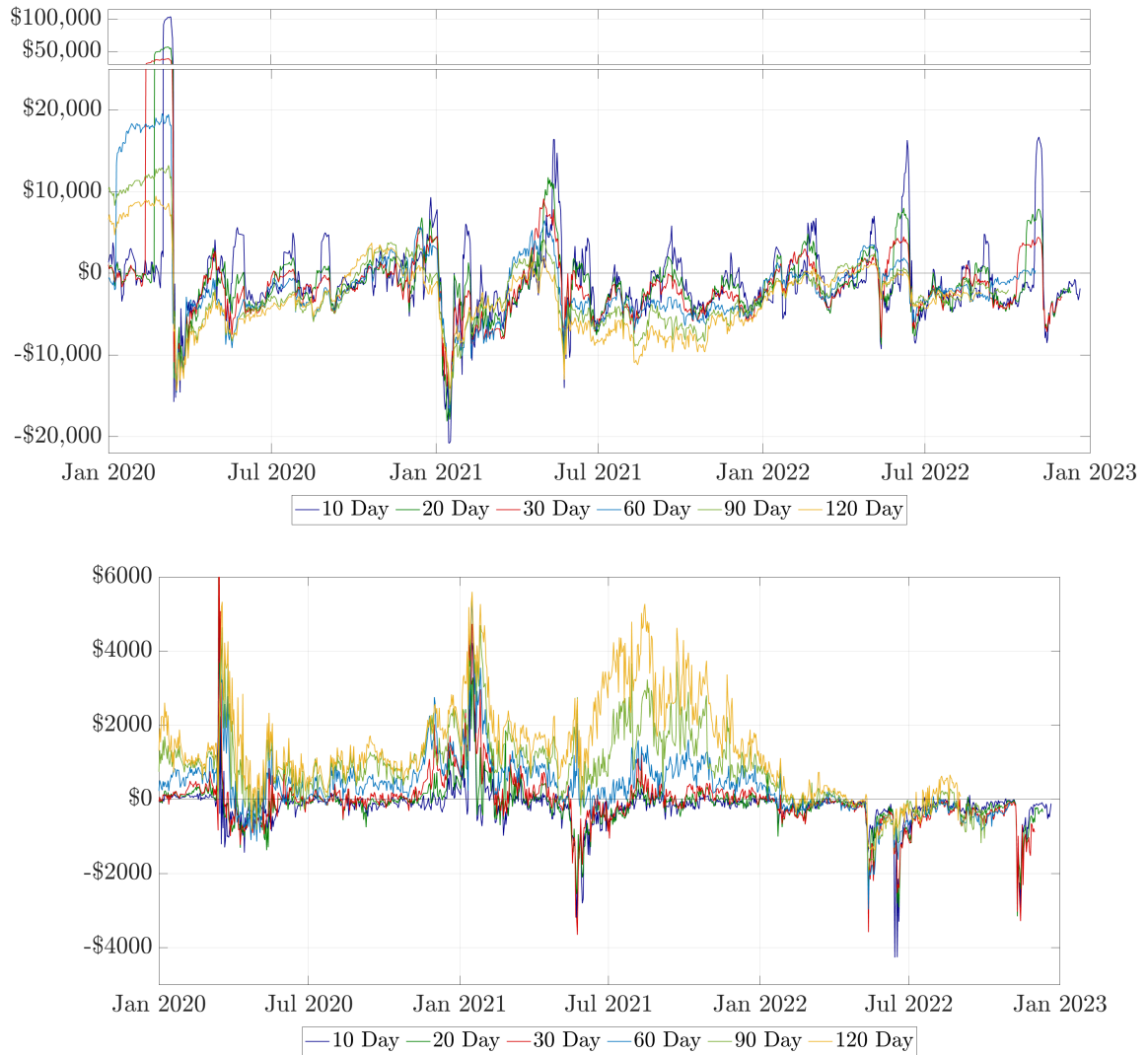
6.4 Determinants of the Bitcoin Volatility Index

This section examines the co-movement of bitcoin variance and its risk premium with more traditional assets viz. the S&P 500 index, crude oil, gold and the USD/EUR foreign exchange rate. This research is an intriguing point for both probationers and scholars. First, understanding bitcoin's implied volatility makes it imperative to understand its dynamics and its potential influence on broader financial stability. Secondly, examining correlations between bitcoin and established asset classes can provide valuable insights into the evolving role of cryptocurrencies within diversified portfolios. Thirdly, as cryptocurrency adoption grows, it becomes more and more important to understand how bitcoin's volatility interacts with more conventional assets, shedding light on potential hedging opportunities or systemic risks. Such an analysis not only enhances the comprehension of cross-asset relationships but also informs risk management. We download 30-day CBOE implied volatility indices for these assets from the CBOE website as follows: for the S&P 500 the VIX; for oil (.USO) the OVX; for gold (.GLD) the GVZ; and for the EURUSD rate (.FXE), i.e. the price of a EUR in USD, the EVZ.[15] We also compare their volatilities in the physical measure, using daily data on the underlying of the options used to compute the volatility indices. The underlying asset data are downloaded from Refinitive and we calculate the RV on a 30 calendar days basis, considering only the trading days that are common to all assets. All prices are taken at the time of the CBOE exchange daily close. Finally, we use the respective volatility indices as a fair-value variance swap rates to compare the VRP on all five assets from January 2020 until 2023.

Figure 6.6 illustrates the 30-day forward looking realised volatility calculated using equation (6.4, upper plot) and the volatility indices for all five assets (bottom graph) between January 2020 to January 2023. The exchange rate exhibits relatively low RV, barely surpassing 9%, even amidst the pandemic. This is unsurprising, as the event affected all currencies and states. Nonetheless, the impact of the COVID-19 outbreak on the other assets is apparent: all assets record their highest RV during March 2020, ranging from 40% (gold) to 170%/180% (oil/bitcoin). Evidently, bitcoin has the highest overall RV, although the oil RV (150%) surpasses bitcoin RV (75%) for roughly two months in April 2020 and subsequently in October 2020 (7%), March and September 2022 (10% each). Between November 2020 and February 2022, the bitcoin and oil RV appear to develop in opposing directions, indicating a negative correlation. The S&P 500 realised volatility fluctuates between 8% and 90% (30% excluding COVID), matching bitcoin's RV on only three brief occasions of approximately five days and appears to be independent from other RV developments. The EURUSD VIX experiences a slight 'jump' from 5% to 9% during COVID, but remains fairly low afterwards till the end of our sample. Both VIX and GVZ oscillate around a similar level of approximately 20%, with their positions interchanging as to which is higher. We expect both bitcoin and oil implied volatility indices to display display similar trends as their RVs. Indeed, bitcoin and oil

---

[15] We would consider macroeconomic factors such as growth or inflation but these are not monitored at sufficiently high frequency to be included in our analysis.

stand apart, developing at significantly higher average levels, i.e. at no point do we find any other index surpassing the BVIN or OVZ. Similar to the RV, the bitcoin VIX surpasses the oil counterpart during most periods. In fact, only during COVID and shortly afterwards does the OVZ exceed the BVIN. Notably, in April 2020, a barrel of oil was traded at about -$40, which led the OVZ jump to over 320%. Beside the absolute magnitudes, all volatility indices appear to evolve similarly, i.e. a sudden upward jump and a correction in a mean-reverting correction afterwards, only exception being bitcoin during January 2021.

Figure 6.6: Realised Volatility and Volatility Indices for different Assets

Forward-looking 30-calendar-day realised volatility of bitcoin (red), equities (S&P 500, blue), oil (grey), gold and FX (EURUSD, green) calculated using (6.4) (upper plot) and the volatility indices of the assets downloaded directly at the CBOE or through Refinitive. Each time series here is constructed at the trading day frequency between January 2020 to 2023 and volatilities are presented in percentage points.
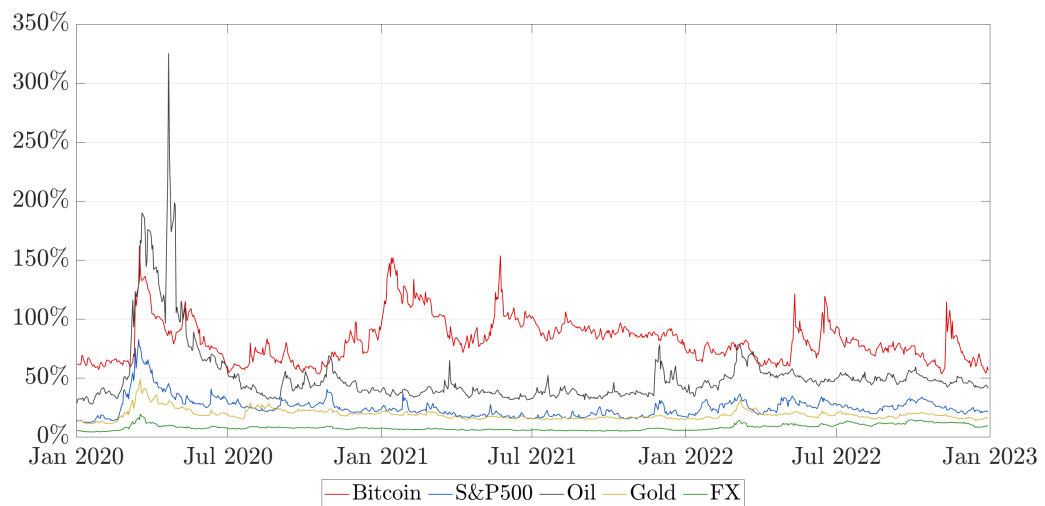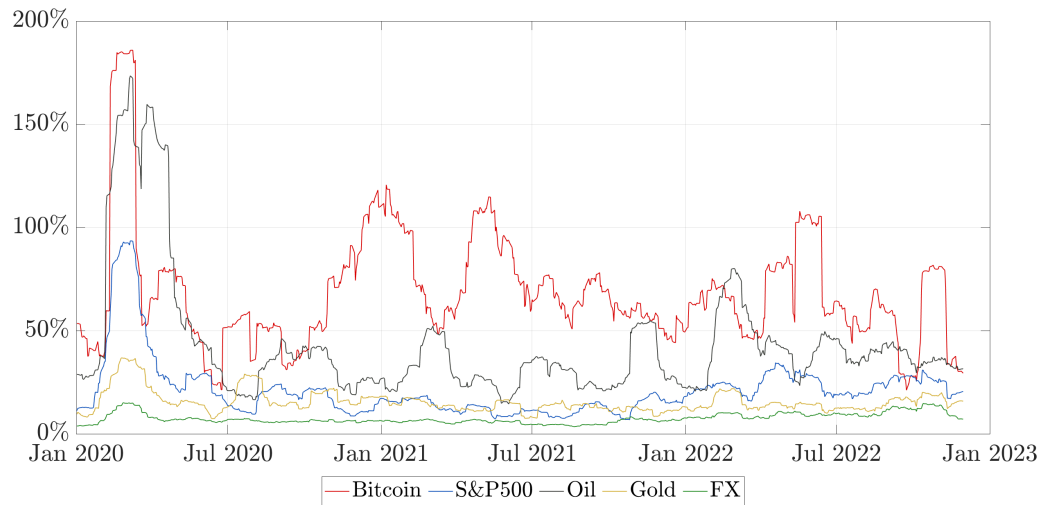
Next, we examine the resulting risk premia for 30 calendar day variance swaps across the five asset classes. Figure 6.7 compares the VRP time series for the respective assets. We retain the same colors for each asset class, but employ distinct notional amounts for different assets when calculating the variance risk premia to enhance clarity. The bitcoin and oil swaps have a notional $N=\$1$, the equity swap has a notional of \$3, gold has $N=\$10$, and finally, $N=\$50$ for the foreign exchange swaps. These notional values serve to scale the VRP to comparable levels in the lower graph, given the disparities in realised volatilities displayed in the upper graph in Figure 6.6. With the exception of bitcoin, all assets adhere to the anticipated pattern of low negative swap payoff with occasional jumps into positive territory following price surges. In fact, across the three-year sample, the S&P 500 VRP is negative in 79.32%, oil VRP in 80.95%, gold VRP in 86.62%, and the EURUSD rate in 64.62% of the time, each with varying magnitudes. In other words, a short position in a variance swap on any of the asset classes under consideration would yield a profit for the majority of the time. Although bitcoin has an comparable distribution of negative values (77.57% of the time), it records considerably lower values than the other assets and displays significantly higher jumps when volatility enters the market. This can be attributed to bitcoin's inherently high volatility level, which is further amplified due to the squared factor of the variance swap.

It is essential to emphasise the March 2020 scenario once more, wherein all assets exhibit a (strong) positive VRP, followed by a substantial (negative) correction. Excluding commodities, throughout the initial quarter of 2020, all assets appear to move in a similar manner, suggesting high correlation. This structure is later disrupted, as the virus's aftermath led investors into presumably safe heavens such as gold. Remarkably, following the COVID-19 crash, bitcoin's VRP remained consistently below zero until the underlying began its rally (December 2020) and reached its (at the time) all-time height in April 2021. Indeed, from April 2021 until June 2022, bitcoin emerged as the 'safest' asset class among those compared concerning the duration of negative risk premia. Over the course of 14 months, bitcoin's VRP jumped into positive territory merely twice and remained negative in 95.11% of the time. In contrast, US equities experienced twice as many positive VRP jumps and remained negative only 84.44% of the time, while oil and FX leapt three times as much as bitcoin, exhibiting a negative VRP in 84% and 54.67%, respectively. Additionally, gold demonstrated three jumps with a negative VRP in 90.18% of the time. In fact, from April 2021 to the end of our sample, bitcoin VRP (80%) remained longer negative than the S&P 500 VRP (76%).

This observation suggests that bitcoin option traders tend to overprice the derivatives, making variance swaps on bitcoin a prime opportunity for writers. Within a hypothetical framework – writing daily variance swaps on each asset class with \$1 notional starting from May 2021 and disregarding transaction costs – bitcoin yields the highest overall (daily average) profit exceeding one million USD (\$2552). Oil ranks second, albeit considerably trailing, with a total (daily) PnL of \$260,350 (\$660), succeeded by equity and gold at \$54,480 (\$138) and \$43,340 (\$110), respectively. A short position on the FX variance swap would be nearly offset.

Figure 6.7: Variance Risk Premia for different Assets

The 30-calendar-day VRP for bitcoin, equities (S&P 500), oil, gold and FX (EURUSD). We keep the same colours and use a notional of N=$1 for bitcoin and oil, N=$3 for equity, N=$10 for gold and N=$50 for FX to depict the VRP more clearly. For the sake of readability, we break the y-axis between -$39,000 to -$20,000; and between $15,000 and $25,000. All calculations are on a daily basis, using the CBOE's closing time. Each time series is constructed at the trading day frequency between January 2020 to 2023.
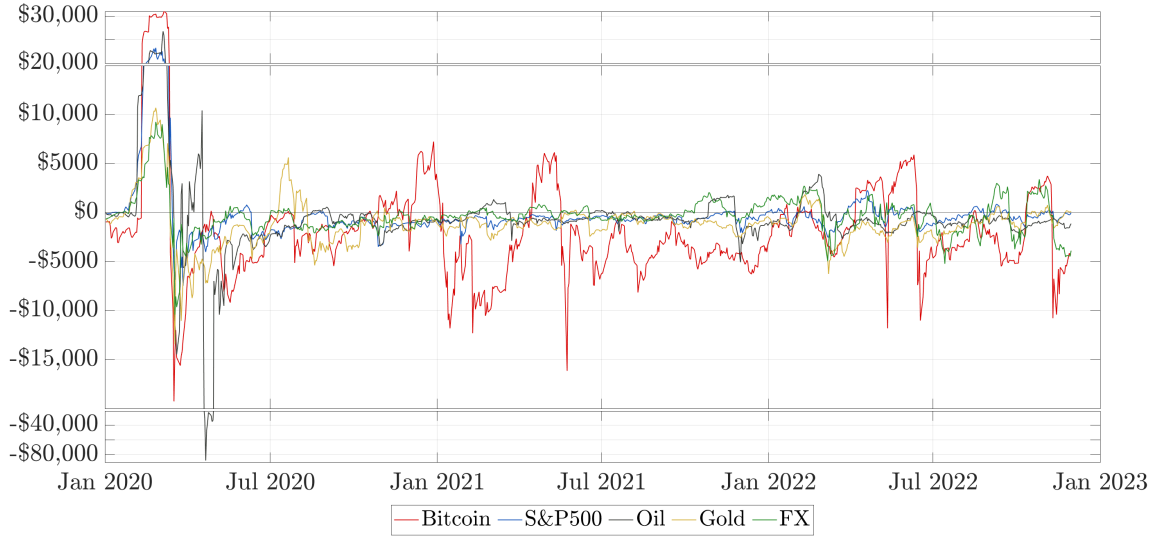
Certain observations from Figures 6.6 and 6.7 are supported by the correlation matrices reported in Table 6.1. It shows pairwise correlations among daily changes in the three different variance matrices depicted in the graphs. The first matrix displays correlations between implied volatility indices, the middle matrix outlines correlations between the RVs of the assets, and the bottom matrix reports correlations between each asset's VRP. Generally, the highest correlation appear among RVs, as shown in the middle matrix of the table and seen in Figure 6.6. To some extent, this can be attributed to the pandemic's impact on the market. It is important to note that the RV does not provide any information on the direction of the volatility. The high (positive) correlation between gold and equity, or gold and FX, is most likely attributed to the leverage effect of equities and safe heaven properties of gold. That is, when stock prices decline (RV↑), investors tend to allocate their capital into commodities, which subsequently elevates their prices (RV↑). Particular emphasis should be placed on the FX column. Note that we consider the EURUSD rate, i.e. the USD value of €1. Consequently, an increase in the EUR is equivalent to the decrease in USD, implying that the purchasing power of $1 diminishes. This factor might explain the (strong) positive correlation between FX and other asset classes' RV.We find the highest correlation of bitcoin RV with equities, which is less surprising given the increasing correlation between bitcoin and equity indices.[16]

---

[16]For example, this Nasdaq article highlights the changes in S&P 500 and bitcoin correlation over recent years and points out an increasing (> 0.6) correlation between bitcoin and US equity indices. Accessed on 01 May 2023.

Volatility indices exhibit lower degree of correlation, with a few notable exceptions concerning equity and gold. These strong correlations signify contagion in the expectations of options traders. For instance, when the S&P 500 index falls, the VIX increases (leverage effect), causing investors to rush into gold, and subsequently raising its price. Traders price this panic-driven behaviour into the options, which increases the GVZ and thus the (strong) positive correlation between the indices. By the same token, we can argue for the positive correlation of gold and FX. Assuming high future volatility for the EURUSD rate, it would most likely drive holder of either side of the currency towards gold, consequently increase both implied volatility indices (Lucey and Tully, 2006). The bottom part of Table 6.1 reveals rather low correlations between changes in the bitcoin VRP and changes in the VRP for all other assets. Paradoxically, despite bitcoin's potential classification as a currency, it demonstrates a very low but significant correlation with the EURUSD exchange rate VRP, making it an optimal diversification choice for any currency portfolio.

Table 6.1: Correlation Matrices

Pearson correlation between daily changes in volatility measures and risk premia in bitcoin, oil, equities (S&P 500), gold and FX (EURUSD). The sample covers all trading days within the period from January 2020 to Janaury 2023 and has a sample size of 761 trading days. We display the statistical significance of the correlation at the 1%, 5% and 10% level with one, two and three asterisk, respectively.

	Bitcoin	**Equity**	**Oil**	**Gold**	**FX**
Volatility Index					
Bitcoin	1	0.147^{***}	0.021	0.115^{***}	-0.026
Equity		1	0.263^{***}	0.356^{***}	0.342^{***}
Oil			1	0.198^{***}	0.244^{***}
Gold				1	0.442^{***}
FX					1
Realised Volatility					
Bitcoin	1	0.342^{***}	0.060	0.221^{***}	0.133^{***}
Equity		1	0.339^{***}	0.237^{***}	0.238^{***}
Oil			1	0.066^{*}	0.145^{***}
Gold				1	0.281^{***}
FX					1
Variance Risk Premium					
Bitcoin	1	0.168^{***}	0.014	0.292^{***}	0.061^{*}
Equity		1	0.153^{***}	0.325^{***}	0.196^{***}
Oil			1	0.082^{**}	0.150^{***}
Gold				1	0.320^{***}
FX					1

We want to investigate the determinants of the bitcoin implied volatility index. For this, we perform a linear regression of the BVIN time series using the other assets as regressors to evaluate to the level to which other assets influence the bitcoin volatility index. That is, we consider:

$$\text{BVIN}_t = \alpha_0 + \alpha_1\text{VIX}_t + \alpha_2\text{OVX}_t + \alpha_3\text{GVZ}_t + \alpha_4\text{EVZ}_t + \epsilon_t,$$

where $\boldsymbol{\alpha} = (\alpha_0, ..., \alpha_4)$ denotes the coefficient vector for the respective volatility indices. We assure stationarity of each implied volatility index time series.[17] Table 6.2 summarises the results of the ordinary least square regression. The OLS model has two statistically significant regressors (GVZ and EVZ) which impact the dependent variable (BVIN), while the other two regressors do not appear to have a significant effect. Surprisingly, equity and oil VIX loose their significant impact on the BVIN. The variance inflation factor (VIF) shows no multicollinearity between the predictors but we find clear suppression effects. That is, we perform individual linear regressions and find that GVZ changes from statistically non-significant in a single regression to significant in a multiple regression. Similar, the VIX becomes statistically non-significant in the multiple regression while being significant in the single. [18] The model as a whole is statistically significant, but only explains approximately 12.1% of the variation in the dependent variable. These findings align with the correlation matrices of Table 6.1 and indicate that only gold (positive impact) and currencies (negative impact) influence the bitcoin implied volatility significantly.

The rationale behind the negative GVZ impact on the BVIN could be attributed to currency fluctuation and the interplay of BTC/USD/EUR. In currency markets, a jump in either direction generates considerable volatility, as FX-rates are considered low-volatility assets. Consequently, the leverage effect for FX-rates may not be as pronounced or consistent as it is in equity markets. Thus, a declining or small GVZ indicates a fairly stable relationship between EUR and USD over the next calendar month. During periods of heightened uncertainty, particularly the two years following the pandemic, the general demand for USD increases as it is commonly perceived as the world's reserve currency. In such instances, we see a leverage effect for currencies, wherein option trader anticipate an increased demand of USD – equivalent to a decrease in the EURUSD rate – and price this expectation into options, leading to a higher GVZ. As the USD value of euro depreciates, the purchasing power of the dollar increases, making it (relatively) more affordable to buy bitcoin options. This (relative) price reduction of the options influences implied volatilities, ultimately decreasing the index. We would expect similar, if not identical, results for other currencies, such as GBP or JPY, as G10 central banks have nearly identical adjustments to their recent interest rate policies.

---

[17]In fact, we try different manipulation techniques, e.g. integrating the data or performing the regression on the log, but none show better results. Additionally, we change the depending variable to any of the other volatility indices to investigate to what extend the BVIN influences the other volatility indices but we do not find any more significant results other than the ones we present here.

[18]These findings confirm Badshah et al. (2013), who perform a similar analysis (excluding bitcoin) and find bi-directional spillovers only for gold and the FX rate.

Hence, no isolated event has led to one currency possessing significant more purchasing power than another.[19] Another explanation might be personal preferences of investors. As the GVZ increases, investors might consider alternative assets like bitcoin to be more attractive regarding their risk-reward profile and consider bitcoin safe(er) in a sense of stability.

An explanation for the positive impact of the GVZ on the BVIN might be bitcoin's diversification potential (Platanakis and Urquhart, 2020), or its promise of an inflation hedge. In the event that fiat currencies loose value, as recently seen with the Turkish Lira or Argentinian Peso, investors tend to rush into commodities like gold or digital commodities, like bitcoin. Another plausible explanation might be due to common investor behaviour. Both bitcoin and gold are often compared due to their shared characteristics, i.e. finite resources and potential stores of value. As such, they might share common drivers such as macroeconomic events of fluctuation in the USD. Some even consider both to be safe havens, although disputable. Hence, the positive relationship might reflect the behaviour of investors trading both assets.

Table 6.2: OLS Regression Results

Summary statistics of an ordinary least square regression on the BVIN. We display the number of observation, degree of freedom of the residuals and model, (adjusted) R^2 and F-Test statistic, among other performance criteria. Furthermore, we show the α coefficient, standard deviation, t-test value and the p-value for the four regressors VIX, OVX, GVZ and EVZ.

Dep. Variable:	BVIN	R-squared:	0.121
Model:	OLS	Adj. R-squared:	0.117
Method:	Least Squares	F-statistic:	26.08
No. Observations:	761	Prob (F-statistic):	2.80e-20
Df Residuals:	756	Log-Likelihood:	-3255.4
Df Model:	4	AIC:	6521.
Covariance Type:	nonrobust	BIC:	6544.

	coef	std err	t	P> |t|	[0.025	0.975]
const	69.638	3.054	22.803	0.0	63.643	75.633
VIX	0.141	0.154	0.917	0.359	-0.161	0.443
OVX	0.027	0.034	0.8	0.424	-0.04	0.094
GVZ	1.346	0.256	5.265	0.0	0.844	1.848
EVZ	-2.23	0.326	-6.847	0.0	-2.869	-1.59

Omnibus:	60.987	Durbin-Watson:	0.124
Prob(Omnibus):	0.0	Jarque-Bera (JB):	78.251
Skew:	0.67	Prob(JB):	1.02e-17
Kurtosis:	3.819	Cond. No.	323

---

[19]However the CBOE does not provide a volatility index on these currencies and we omit them from our research.

6.5 Concluding Remarks

Throughout the course of this chapter, we construct a new set of volatility indices which we call the crypto 'investor fear gauge', after the classic paper by Whaley (2000) that introduces the equity investor equivalent. The bitcoin market structure differs from markets for other assets because there are literally hundreds of spot exchanges and numerous derivatives exchanges, almost all of them are unregulated and they trade 24/7 on every day of the year. As such, the bitcoin price moves very much more quickly than those for standard asset classes. Thus, when designing volatility indices for bitcoin, in addition to the standard curve of one month and longer we examine the empirical behaviour of bitcoin volatility indices that are constructed using the CBOE methodology at short (10 days), mid (20 days), long (30 days) and ultra-long (60, 90 & 120 days) terms. We compare the indices with those based on an alternative methodology which represents a fair value for the average sum of squared price increments and which has no jump risk, unlike the CBOE formula

We use the k-day bitcoin VIX as the fair-value of an k-day variance swap, compare this with the k-day realised volatility and hence examine the empirical properties of the bitcoin variance risk premium. The premia have some features in common with the variance risk premia on other assets. While the swap payoff is generally negative with occasional jumps, the magnitude of bitcoin VRP across all maturities is much more pronounced and jumps occur more frequent for short-term swaps. However, it is not clear that they are closely correlated with the variance risk premia on other assets. Bitcoin itself has been regarded as a safe-haven asset that offers considerable diversification potential for traditional assets and we ask the question whether this is also true of for variance?

To address this question we compare our 30-day bitcoin VIX with the CBOE volatility indices for the S&P 500 (the original VIX), crude oil, gold and the EURUSD exchange rate. We examine the relationships between all five assets in terms of their realised volatility, implied volatility and variance risk premia. Bitcoin itself and also its variance has been behaving very similarly to traditional assets. Its implied volatility and realised volatility shows positive correlation with all assets except the FX rate. Remarkably, bitcoin emerges as the safest asset in our portfolio, as its variance risk premium sustains a negative value longer than any other asset over an extended duration of 14 months. Consequently, any variance swap writer would have reaped great profits from writing variance swaps on bitcoin. In fact, by the end of our data sample in January 2023, maintaining a short position in the bitcoin variance risk premium proved more lucrative than the equity variance risk premium. Finally, we investigate the relationship between our volatility index and the indices of other asset classes to discern their influence on the BVIN. Only gold (positive) and foreign exchange rates (negative) exhibit a significant relationship with the BVIN. In summary, it remains challenging to attribute diversification potential to Bitcoin or its variance, at least within the prevailing market conditions. However, it does present an exceptional opportunity for any writer of variance swaps.

CHAPTER 7

IMPLIED VOLATILITY DYNAMICS

Motivation: The option implied volatility is essential for market participants and traders. It serves as a key metric for future market volatility and reflects the belief of sophisticated traders about future returns of the underlying. Moreover, the implied volatility acts as the standard pricing unit for options among traders and plays a significant role in risk management. Naturally, the ability to forecast and control the dynamics of future changes in the implied volatility is of utmost importance for pricing and hedging options, as well as for trading systematic volatility. Practitioners and academics still try to exploit the predictability of the surface.

Summary: We model and predict the dynamics of the bitcoin implied volatility surface. To accomplish this, we employ a robust cross-sectional ordinary least square to fit intra-day surface data to a second-order Taylor expansion and generate a time series of model parameters. Subsequently, we use a vector-autoregressive and long-short-term memory model to analyse and forecast the time series and evaluate the predictions based on a set of statistical measures, e.g. squared, directional and absolute error. To evaluate the economical impact, we assess the profitability of two potential trading strategies derived from our predictions and benchmark their performance against that of a naive model. Although we are able to predict the surface accurately for some moneyness-maturities and leverage this to create profitable trading strategies, real world implications such as transaction costs and spreads nullify the profit.
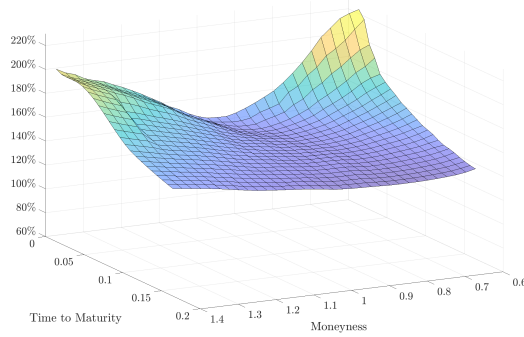
7.1 Modelling the Implied Volatility Surface

The Black and Scholes (1973) assumption of constant volatility throughout the entire options chain has been refuted by numerous studies in the financial literature. In fact, the implied volatility depends strongly, but not exclusively, on the option's strike level and its time to maturity. On a more general level, even on the asset class itself. These variations in implied volatility across different strike level and TTM result in the implied volatility surface. For equity index options, e.g. S&P 500 index options, the surface structure has been researched extensively and is well understood. Particularly, the so-called 'leverage effect' describes a (strong) inverse relationship between ATM implied volatilities and the underlying asset, see for example Haugen et al. (1991) or Bouchaud et al. (2001). During periods of low volatility, the implied volatility's term structure displays an concave structure, increasing with maturity, which is inverted in high volatility regimes, and the entire surface shifts upwards. The strike-dependency results in the famous volatility skew which resembles an asymmetric smirk or smile, i.e. the OTM put exceeds OTM call IV, which tends to flatten out for longer maturities.

Unfortunately, these simplified rules of thumb do not hold entirely for the bitcoin implied volatility surfaces. Figure 7.1 provides an exemplary illustration of two bitcoin implied volatility surfaces backed out from traded option prices on Deribit on 15 January 2021 and 17 March 2022. The surfaces cover moneyness ranging from 0.6 to 1.4 and maturities between five days and two months. For clarity, we interpolate and extrapolate in an arbitrage-free way considering only ATM and OTM options due to higher liquidity. Note that, like options on other asset classes, bitcoin implied volatilities display a clear strike-dependency, resulting in an asymmetric skew that holds across all maturities. The convex 'U'-shape of the skew is most prominent for short-term maturities and flattens as TTM increases. Considering the implied volatility skew of equity index options as a function of moneyness, it typically displays a higher IV for OTM puts than calls with same delta due to the negative correlation between underlying and volatility. Hence, put options provide an excellent insurance against falling index values and traders are willing to pay more. The March 2022 bitcoin surface illustrates a similar smile pattern; however, this observation is not consistent with the January observation. Specifically, while the skew keeps its smile structure, the IVs for OTM calls exceed their put counterparts. The short-term deep OTM ($m = 1.4$) call option in January 2021 depicts a 20 percentage point surplus compared to its put counterpart ($m=0.6$). As we have seen in Section 5.2, the leverage effect does not hold for bitcoin which indicates either an extremely bullish trend or a positive correlation between underlying returns and volatility, a phenomenon hardly seen for equity options. Furthermore, we observe a substantial change in the overall surface level between the two dates despite similar underlying values (\$39,500 in January 2021 and \$41,000 in March 2022). This characteristic, however, is not unique to bitcoin as it is well known that IVS dynamically changes its shape and level depending on traders' beliefs of future returns. Overall, the base level of the surface has changed from 120% to 80% due to different prevailing market regimes.

Figure 7.1: Bitcoin Implied Volatility Surfaces

Implied volatilities backed from traded bitcoin options on Deribit on 17 March 2022 and 15 January 2021, respectively. The underlying prices on these days were $41,000 and $39,500. We illustrate options with moneyness ranging from 40% below (put) and above (call) the current underlying value. The maturities range from five days to two months, covering more than 95% of the overall traded volume. We interpolate the surface to finer grids in an arbitrage-free way.

(a) 15 January 2021
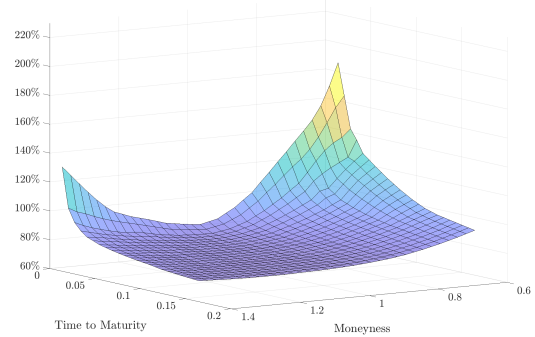
(b) 17 March 2022

Modelling the implied volatility surface and its dynamics poses significant challenges, let alone predicting it. A variety of models are capable of explaining the observed behavior documented above. However, sudden shifts in the surface, variations in shape, and uneven changes in the smirk, particularly for the tails, make capturing it a difficult task. In this chapter, we aim to capture and predict the dynamics of the bitcoin IVS and examine their behavior. Our approach is twofold: we first consider a parametric representation of the surface, and subsequently model the time series of the parameters. Intriguingly, we are able to predict the surface accurately and leverage this information to execute profitable trading strategies. Nonetheless, transaction costs and the bid-ask spread eliminate our profits.

7.2 Experimental Design

The academic standard of predicting implied volatility is two-fold. First, we capture the key elements driving the surface; and second, we model the time series of these elements. As such, this study examines (i) the accuracy of the chosen model to capture the surface; and (ii), how well the time series model predicts changes.

7.2.1 Implied Volatility Model

Daglish et al. (2007) propose an intuitive approach, where the implied volatility, given as a function of moneyness and time to maturity, is modelled through a second-order Taylor

expansion around the two dependent variables.[1] This approach builds on the earlier implied volatility model proposals of of Rubinstein (1994), Derman and Kani (1994) and Dumas et al. (1998), who emphasise different (multi-dimensional) polynomial functions of quadratic order to express the implied volatility:

$$\sigma_t(m, \tau) = \beta_{0,t} + \beta_{1,t}m + \beta_{2,t}m^2 + \beta_{3,t}\tau + \beta_{5,t}m\tau + \epsilon$$

or

$$\sigma_t(m, \tau) = \beta_{0,t} + \beta_{1,t}m + \beta_{2,t}m^2 + \beta_{3,t}\tau + \beta_{4,t}\tau^2 + \beta_{5,t}m\tau + \epsilon,$$

where $\sigma_t(m, \tau)$ is the implied volatility at time t of an (OTM) option with moneyness m and TTM τ, and ϵ describes the noise or error term.[2] These polynomial models are well-suited to capture the instantaneous implied volatility across both the moneyness and maturity dimension, whilst providing an intuitive interpretation of each beta coefficient. At the same time, these models offer a simple and easy-to-implement approach which is often preferred by practitioners and tend to level the results created by more sophisticated models. Furthermore, unlike stochastic volatility models, this representation requires no additional assumptions.

Each parameter corresponds to an observable surface characteristic: The constant, β_0, captures the overall level of the surface; the parameter β_1 captures the slope of the implied volatility smile with respect to the moneyness, while β_2 reflects its curvature. By the same token, β_3 reflects the slope of the term structure and β_4 its curvature. Any possible synergy between time to maturity and moneyness is seized in β_5. However, despite providing excellent surface fits, this approach presents slightly unstable parameters over time as Dumas et al. (1998) points out but this poses a negligible problem since the time series are stable enough to to be modelled, as Gonçalves and Guidolin (2006) explain in their two-step approach.

In light of the documented advantages and the outstanding success for equity index implied volatility prediction and trading (Bernales and Guidolin (2014)) as well as cross-sectional pricing (Brandt and Wu (2002)), we opt for a deterministic IVS representation of Dumas et al.

---

[1]Recall that the aim of this study is not to price or hedge bitcoin options across the surface accurately, but rather predict their implied volatility a step ahead. Alternatives involve various filtering methods and principal component approaches in which surface data is reduced to a set of driving factors. (Cont and Da Fonseca, 2002) first smoothing and then model the S&P500 or various foreign-exchange (FX) surfaces (Beer and Fink, 2019) using a Nadaraya-Watson estimator with a Gaussian kernel. However, these type of models require additional steps in form of a principal component analysis (PCA) to generate a time series. Consequently, predictions are less intuitive for traders and the multi-step process gives little incentive to implement these models in a practical environment. Furthermore, a PCA approach yields best results if correlation between parties is high, i.e. high correlation between implied volatilities for different strike and maturity, which is not the case for bitcoin ATM/OTM IVs, as seen in Chapter 5. Fengler et al. (2007) propose a semiparametric factor model in which the use a the quartic kernel, similar to a Kalman filter, to estimate three parameters on a three year daily DAX IVS dataset. Similar to our approach, they use a VAR model to forecast their parameter yielding accurate surface forecasts, see Audrino and Colangelo (2009) for a summary.

[2]We excluded two more models since these reflect the BSM assumption of a constant volatility and the a model which captures only the smirk, similar to the second order polynomial representation of the IV smile proposed by Lewis (2000). However, these two models are not suitable to model the surface. First, the surface is never constant, second, the absence of a maturity factor makes it impossible to obtain feasible values in the first place.

(1998). However, it is important to note that the two model mentioned earlier form just a baseline. Hence, we explore various data manipulation techniques that could potentially lead to an improved overall fit. Pena et al. (1999) compare and extend the structural forms of Dumas et al. (1998) to different implied volatility (smile) models. In addition to the existing models, the authors introduce alternative parametric frameworks using log moneyness for the explanatory variables and conclude that this data transformation yields a better fit compared to the base models. Other authors (Tompkins and D'Ecclesia, 2006) advocate the use of time-adjusted moneyness, $M(m, \tau) = \frac{\ln m}{\sqrt{\tau}}$, rather than standard moneyness. Although this approach is intended to normalize moneyness across all maturities, it could potentially affect the explanatory power of the moneyness as it depends on another variable. Furthermore, both Gonçalves and Guidolin (2006) and Bernales and Guidolin (2014) assume the relationship expressed in the previous models holds for the log surface. We consider all feasible deterministic models and decide on the best for our further research. For this, we compare numerous variations of the base model documented above, taking into account goodness-of-fit evaluated based on various criteria. We classify the variations into three main categories: *base*, *log* and *time-adjusted moneyness (tam)* models, and test the fit with different numbers of parameters within each category. In total, we test the following models:[3]

$$\sigma_t(m, \tau) = \beta_{0,t} + \beta_{1,t}m + \beta_{2,t}m^2 + \beta_{3,t}\tau + \beta_{5,t}m\tau + \epsilon \qquad \text{(Base I)} \quad (7.1)$$

$$\sigma_t(m, \tau) = \beta_{0,t} + \beta_{1,t}m + \beta_{2,t}m^2 + \beta_{3,t}\tau + \beta_{4,t}\tau^2 + \beta_{5,t}m\tau + \epsilon \qquad \text{(Base II)} \quad (7.2)$$

$$\sigma_t(m, \tau) = \beta_{0,t} + \beta_{1,t}\ln m + \beta_{2,t}\ln m^2 + \beta_{3,t}\tau + \beta_{5,t}\ln m \ \tau + \epsilon \qquad \text{(Log I)} \quad (7.3)$$

$$\sigma_t(m, \tau) = \beta_{0,t} + \beta_{1,t}\ln m + \beta_{2,t}\ln m^2 + \beta_{3,t}\tau + \beta_{4,t}\tau^2 + \beta_{5,t}\ln m \ \tau + \epsilon \qquad \text{(Log II)} \quad (7.4)$$

$$\ln \sigma_t(m, \tau) = \beta_{0,t} + \beta_{1,t}M(m, \tau) + \beta_{2,t}M(m, \tau)^2 + \beta_{3,t}\tau + \beta_{5,t}M(m, \tau)\tau + \epsilon \qquad \text{(Tam I)} \quad (7.5)$$

$$\ln \sigma_t(m, \tau) = \beta_{0,t} + \beta_{1,t}M(m, \tau) + \beta_{2,t}M(m, \tau)^2 + \beta_{3,t}\tau + \beta_{4,t}\tau^2 + \epsilon. \qquad \text{(Tam II)} \quad (7.6)$$

$$\ln \sigma_t(m, \tau) = \beta_{0,t} + \beta_{1,t}M(m, \tau) + \beta_{2,t}M(m, \tau)^2 + \beta_{3,t}\tau + \beta_{4,t}\tau^2 + \beta_{5,t}M(m, \tau)\tau + \epsilon \qquad \text{(Tam III)} \quad (7.7)$$

In the following paragraph, we outline the parameter estimation method we use. While a simple cross-sectional ordinary least squares regression is the intuitive approach to estimate the beta coefficients, Hentschel (2003) cautions that measurement errors such as non-synchronous underlying prices or tick sizes could lead to heteroskedasticity in the residuals or autocorrelated errors of the OLS. Such violations of the OLS assumptions may result in false or biased conclusions.[4] To mitigate this risk, we employ a robust cross-sectional OLS and use

---

[3]Note that there exist theoretically infinite permutations of the base models documented above. Each extension of this would probably yield a better results. However, we need to consider a feasible trade-off between further improvements and overcomplicating the problem. We decide to consider only the following as these has been proven effective for alternative asset classes in the past. Furthermore, the betas and error notation is the same across all models. This, however, is a manipulative use of notation and should not imply that the betas are identical across the models. We use this notation for the sake of clarity and shall distinguish more precisely at a later stage if needed.

[4]The author proposes an alternative approach to fit the deterministic model using the generalized least square (GLS) method which takes into account the covariances of the residuals.

the standard Newey and West (1994) method to correct standard errors for heteroskedasticity, autocorrelation and other misspecification of the data. Furthermore, we validate the stationarity of the dependent variable data by performing an Augmented Dickey-Fuller (ADF) test on the time series of constant-moneyness and constant-maturity (log) implied volatilities.[5]

At any time step t we estimate the vector of coefficients $\boldsymbol{\beta}_t = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)^\mathsf{T}$ and $\boldsymbol{\beta}_t = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5)^\mathsf{T}$, respectively, using a robust cross-sectional OLS on all suitable implied volatilities. For the sake of clarity and readability, we omit the time-dependent subscript and model-dependent superscript for each single beta coefficient. First, we filter out all options that violate the no-arbitrage, no-trading or spread constrain explained in Section 3.1. While other researchers often filter out short-term options, we retain these in our dataset as long as they satisfy the no-arbitrage and trading conditions previously mentioned. Indeed, the interest in bitcoin options is heavily concentrated on short-term options up to ten days, as seen in Section 3.2. However, we filter out ultra-short-term options, i.e. those with less than three days to maturity. We do this because (i) ultra-short-term options generally trade only one ATM strike, thus not adding much information to the surface shape; and (ii), modeling the implied volatility of these ultra-short maturities proves immensely challenging, as the term structure shows 'bumps' in the implied volatility.[6] We then perform regressions based on the seven models presented above, i.e. using various manipulated market implied volatilities, to determine which yield optimal estimates for $\boldsymbol{\beta}$. To ensure viable results, we only conduct the regression if the sample size exceeds 50 suitable options. These regressions are conducted on market data snapshots captured every eight hours from August 2019 to August 2022, resulting in over 3200 observations and more than 1.5 million option prices. This process produces a time series of beta vectors, or multiple time series of individual beta parameters.[7] Additionally, we provide various statistical measures to evaluate the model fit.

Table 7.1 presents a comparison of various model selection criteria, including the average Akaike information criterion (AIC), the Bayesian information criterion (BIC), (adjusted) R^2 with a special emphasis the root mean square error for each of the tested models over all regressions.[8] We concur with the findings of Pena et al. (1999) that the log model provides a

---

[5]We perform an augmented Dickey-Fuller test on the (log) implied volatilities time series for different constant moneyness and maturity options, interpolated in an arbitrage-free way. Independent of the moneyness or maturity, we are able to rejecting the null hypothesis of an unit root existence at 1% significance, hence assuming the data to be stationary.

[6]This phenomena is not exclusive visible for bitcoin and seen in ever asset class (Medvedev and Scaillet, 2007). The delta-gamma-vega hedging costs of these options are immense, resulting in a very sensible area of the surface. Indeed, as an example, on October 4, 2022 the day-ahead ATM option had a IV of 59.98%, the 2 Days to Expiration (DTE) 60.73%, the 3 DTE 63.71%, the 10 DTE 63.15% 17 DTE 63.05%, the 24 DTE 62.91% and the 52 DTE 65.69%. This 'bump' in the IV term structure, albeit small, is quite common for these ultra-short term options across all asset classes.

[7]Contrary to the two previous chapters, we include a much wider dataset. Note that we considered in Chapters 5 and 6 only volatility skews for certain maturities and required a wide range of strikes. While this was not given prior to 2020, we are confident that the options data is sufficient enough to draw conclusions on the entire implied volatility surface. Additionally, a larger dataset will be beneficial for further research in this chapter. Overall, the average number of contracts per regression was 117. The lowest number was 50 and the maximum 245.

[8]To evaluate model performance and account for real-world conditions, we divide our data into three sets:

superior fit compared to the base form. However, the models with time-adjusted moneyness demonstrate exceptional performance, surpassing both base and log models across all metrics. Among the Tam models, Tam III offers the best fit, largely due to its incorporation of six beta coefficients, compared to the five independent parameters in Tam I and Tam II. Furthermore, the penalty term for the extra parameter does not seem to significantly affect the fit, as suggested by the smaller AIC and BIC values compared to the other Tam models. Based on these results, we opt to exclude the base and log models from further consideration and proceed with the Tam models.

We now evaluate the significance of each beta parameter on the implied volatility, considering the average p-value of a two-sided t-test for each parameter across all regressions. Surprisingly, the β_5 parameter, which is the coefficient capturing synergies between moneyness and maturity, exerts only a slight impact on the implied volatility. This coefficient exhibits a significant impact in only 46% (Tam I) and 56% (Tam III) of cases, respectively, using a standard 10% significance level. Despite this, we choose not to eliminate this model based solely on one less significant parameter. In Tam II, each coefficient demonstrates a substantial impact on the implied volatility fit, as indicated by the 25th percentile of each beta parameter's p-value being below 0.01. This high significance of coefficients leads us to select Tam II as our primary model, even though its fit is second-best compared to Tam III. For the sake of clarity and readability, henceforth we will refer to Tam II (7.6) as 'Model 1' and Tam III (7.7) as 'Model 2'.

Table 7.1: Model Evaluation Criteria

Summary of statistical model evaluation measures. For each model, we present the Akaike information criterion (AIC), Bayesian information criterion (BIC), R^2 as well as adjusted R^2 value and the root mean square error (RSME) as an average over each regression. The cross-sectional regressions are performed considering suitable options on an eight hour frequency coverings a three-year period starting from August 2019.

Criterion	Base I	Base II	Log I	Log II	Tam I	Tam II	Tam III
AIC	-208.76	-219.2	-244.57	-256.58	-350.77	-372.74	-382.91
BIC	-195.63	-203.44	-231.44	-240.83	-337.64	-359.61	-367.15
R^2	0.419	0.475	0.571	0.6178	0.86	0.886	0.898
Adjusted R^2	0.394	0.447	0.553	0.597	0.853	0.88	0.892
RMSE	0.092	0.087	0.079	0.074	0.053	0.047	0.045

To support our decision, we replicate implied volatilities using beta coefficients derived from the two models (7.6) and (7.7), and compare these volatilities with the observed market surface data. Figure 7.2 illustrates the market implied volatility alongside the volatility replicated from the fitted beta coefficients of the two models. The top row of the figure focuses

---

an in-sample (or train) set, a validation set, and an out-of-sample (or test) set. This division is based on a 30:6:5 month ratio and covers the periods August 2019 to January 2022, February 2022 to July 2022, and August 2022 to end December 2022, respectively. We adopt this ratio in line with the recommendation of Hastie et al. (2009), which allows for robust training of potential relationships while accommodating different regimes. For each of the models, we define and calibrate the optimal model parameter using the 73% of the data.

on the implied smile for synthetic 30-day constant maturity, with a strike range between 0.8 (OTM puts) and 1.2 (OTM calls). We interpolate observed implied volatilities using shape-preserving PCHIP splines in an arbitrage-free manner for the market IVs (top left graph), while we calculate the smiles for Model 1 and Model 2, respectively. Both models replicate the general smile and smirk shape, with the transition from negative to positive skew clearly evident in both models, mirroring actual market observations. Overall, the smile replication using either model yields highly satisfactory results and correctly captures the general level of the smile. Furthermore, we examine whether the arbitrage-free constraints hold for the modelled surface and find no violation for any strike level or maturity up to nine months. Given the model structure, the long-term maturities might violate the no-arbitrage constraints. We find that on some occasions this happens for maturities longer than nine months. Therefore, we omit these from further consideration. However, neither model is capable of *perfect* replication, particularly for out-of-the-money tails and extreme long-term structures. Indeed, we find significant discrepancies, particularly for deep OTM options.

Figure 7.2: Market and Model Implied Volatility Comparison

Comparison of daily market and model implied volatility smiles (upper row) and term structure (lower row) over a three-year sample period starting from August 2019. Both the $x-$ and $z-$axis are identical for all six graphs, i.e. the date ($x-$axis) and implied volatility in percentage points ($z-$axis). The $y-$ axis on the upper half is given as moneyness (K/S) while the lower $y-$axis reflect the options' time to maturity in years. The left column depicts the actual market skew and term structure, the middle column presents the implied volatilities calculated using the betas of Model 1 in (7.6) and the last columns illustrates the results obtained using Model 2 from (7.7). The upper half of the figure considers 30-day constant maturity bitcoin options with strike levels 20% below and above the current underlying values. This is due to the high liquidity for wide strike ranges of this particular maturity. We decide to consider only OTM implied volatilities, i.e. puts with $m < 1$ and calls for $m \geq 1$. The lower half illustrates the ATM IV term structure of bitcoin options, ranging between three days to three months till expiry. We decide to focus on the ATM options as these are the most traded strike level s with highest information content. Maturities between one and three months already lack trading volume, hence there is no point in considering longer maturities.
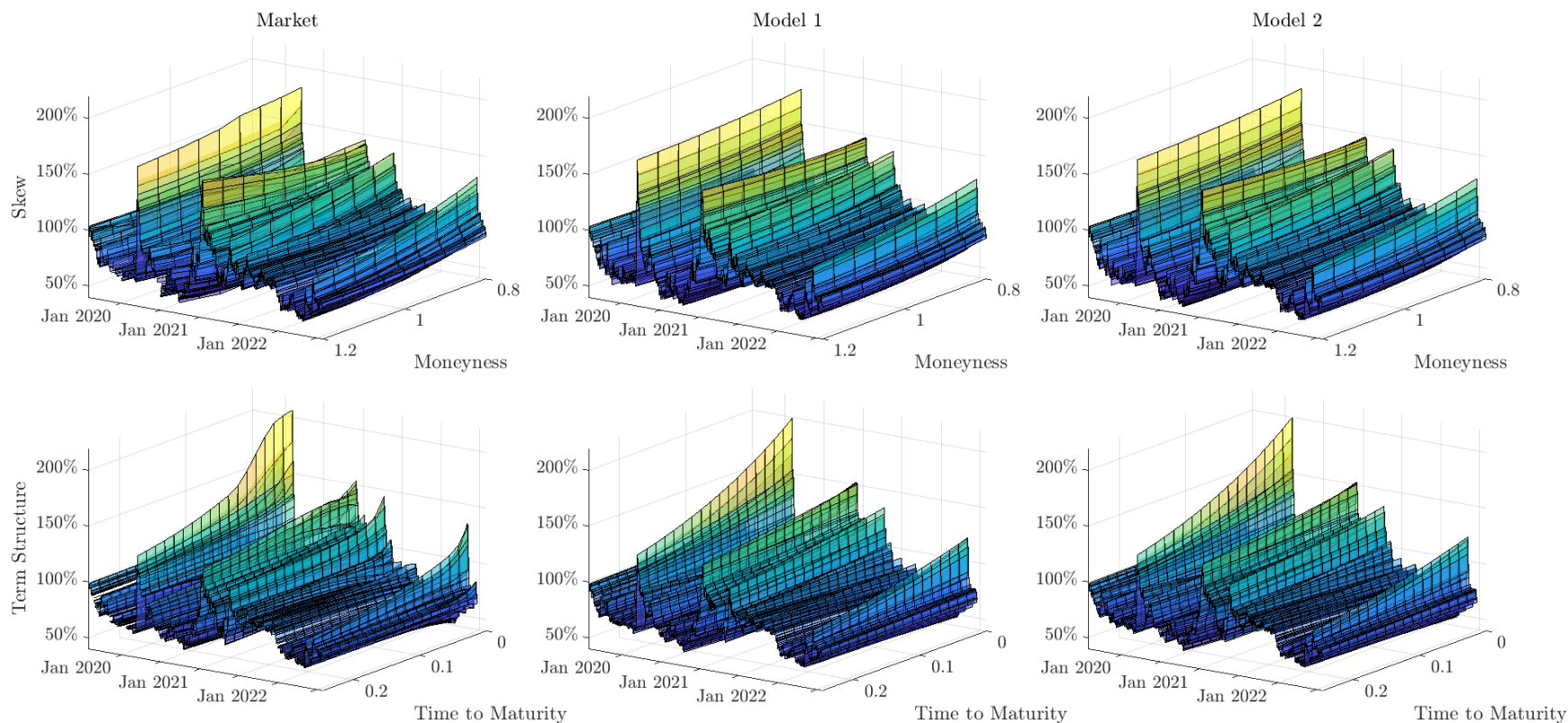
Given that both models are proficient in replicating the surface, we aim to analyse each beta coefficient in more detail. Figure 7.3 illustrates the time series of the four beta coefficients estimated using the robust OLS method from Model 1 in (7.6), depicted in red, during our in-sample period from August 2019 to August 2022 at eight-hour intervals. It presents each beta coefficient on the same time scale, in increasing order from β_0 to β_4. Likewise, Figure 7.4 depicts each beta coefficient estimated using Model 2 in (7.7), represented in blue, on the same time scale and frequency. The first four coefficients of both Model (7.6) and (7.7) capture the same surface characteristics. Consequently, we observe that the time series of each coefficient follow comparable paths, despite not being identical in terms of their values.

Take, for example, the intercept of the regression, denoted by β_0. This coefficient measures the overall surface level. Consequently, we expect the time series of β_0 to resemble the constant maturity ATM implied volatility development in Figure 5.1. For both models, the intercept fluctuates between ± 1 exhibiting a mean-reverting behaviour, with sudden positive jumps following bitcoin price shocks. As we are modelling the log surface, a coefficient value of zero implies a constant volatility of 100%. The parameter generally exhibits negative values throughout most of the sample period, indicating base volatility levels below 100%. Note that modelling the IVS solely through β_0 would imply the BS assumption of constant volatility. Contrary to prior research, we did not detect any significant instability in this parameter over time. In fact, the parameter exhibits no more noise than the ATM implied volatilites.

The next parameter, i.e. β_1, captures the slope of a linear representation of the implied volatility smile for increasing (decreasing) strike levels, similar in concept to Derman and Kani (1994), who model the implied smile as a linear function of the strike level and ATM IV. Analogous to the first beta coefficient, both models show an almost identical path but differ in absolute values. The economic interpretation of this parameter is a general correlation between the strike level and IV. A negative value indicates an increasing IV for decreasing option delta. Throughout the sample period, we would generally expect β_1 to be negative, as deep OTM put options are considered an effective insurance against market crashes and, hence, have a greater IV compared to their OTM call counterparts. However, the alternating behaviour of this coefficient for bitcoin implied volatility indicates a shift in the IV smiles, where traders value a deep OTM call over a put, in line with the findings of Section 5.2. Here again, we find a surprisingly stable development throughout time.

However, isolating this parameter will not yield any useful information as it displays the smile as a straight line. It needs to be considered in combination with β_2, as this quadratic factor gives rise to the familiar smile/smirk structure of the IV. Hence, β_2 depicts a strictly positive development throughout the sample data and assures the convexity of the IV smile, where the magnitude of β_2 influences the intensity of the curvature, i.e. a high (low) beta results in a steep (flat) smile. A common observation after underlying price jumps is that option market makers price this uncertainty in a way that raises the general IV level and flattens the smile across the moneyness. Thus, it is not surprising that β_2 reaches its lowest

values shortly after jumps (March 2020, June 2021 or May 2022) when the smile curvature is minimal. Some authors (Lewis, 2000) stop here and propose a linear combination of β_0, β_1, and β_2 to replicate the IV smile. Overall, the time series depicts a satisfactorily stable trend.

In a similar vein, β_3 and β_4 capture the same IV attributes, but consider the implied volatility as a function of time to maturity. For instance, β_3 captures the co-movement of IV with increasing time to maturity, i.e. an expression of the IV as a linear function of maturity. Note that the longest, currently *actively* traded expiry for bitcoin options does not exceed six months, which makes a comparison to the equity IV term structure rather challenging, where we see maturities of up to two years. In the case of bitcoin, the term structure follows the well-known concave saturation curve, similar to the term structure of interest rates, but is much less pronounced. The difference between short- and long-term ATM IV rarely exceeds five percentage points, except during tranquil periods, when the term structure inverts after crashes. Nevertheless, β_3 can capture the increasing (decreasing) term structure as seen in Figure 7.2. Positive coefficients indicate increasing IV for higher maturities, with clear identification of bitcoin jumps (March 2020, May 2022) and the bubble fear period (Winter 2020/21) as β_3 becomes strongly negative, pointing towards an inverted term structure where short-term IV exceeds long-term IV. Initially, we find a rather unstable development of the parameter which could be a statistical artefact of the relatively low sample period of around 50 options. However, the time series subsequently stabilizes with distinct regime-dependent tendencies. After the COVID crash, the coefficient rarely exceeds the lower limit of -1 and the upper limit of 2. Particularly, during the last year starting from July 2021, it demonstrates a range-bounded development hardly exceeding ± 1, except for May 2022 during the TerraLuna collapse. In contrast to the two coefficients capturing smile dynamics, which have no other restriction but a positive curvature, we expect the term structure betas to be negatively correlated to replicate the bond-type term structure of the implied volatility. Indeed, at first glance, the β_4 time series looks like a mirrored version of β_3 with different absolute values.

Note that up to this point, the beta coefficients exhibit similar patterns, albeit with different absolute values. In Model 2, we introduce an additional coefficient, β_5. By conceptualising the surface in a three-dimensional, xyz-Cartesian coordinate system, where the x-axis represents the option's moneyness, the y-axis represents the time to maturity, and the z-axis represents the implied volatility, Model 1 is capable of capturing movements parallel to the x-axis (β_1, β_2) and the y-axis (β_3, β_4), while Model 2 can capture additional movements on the xy-plane (β_5). Arguably, β_5 plays an important role in surface modelling, as it allows for a relationship between two distinct grid points on the surface that are not parallel to the x- or y-axis, despite only having a significant impact on the overall fit roughly 60% of the time. This coefficient exhibits a range-bounded evolution, uncorrelated with the other four parameters. Specifically, after January 2021, β_5 rarely exceeds a value of around ± 0.15.

It is worth noting that all beta coefficients in both models exhibit greater stability after January 2021. This could be due to several factors, including a maturing options market, an increasing number of tradable options, or newly issued maturity/strike combinations, which create a finer surface grid. Nonetheless, the apparent stability does not justify splitting the dataset, even though restricting the analysis to the second half of the sample produces more efficient model results.[9] These potential improvements do not outweigh the inevitable loss of information contained in 1.5 years of historical surface data, particularly in hindsight of training deep learning models which require an extensive amount of data.

Table 7.2 reports an array of statistics for the estimated coefficients using Model 1 (upper panel) and Model 2 (lower panel) over the past three years using the robust OLS regression. It presents – for each beta – the mean and standard deviation of the time series, the average of a two sided t-test over all regressions as well as the percentage of the time is shows significance at a standard 10% in parentheses. To evaluate the stationarity of the time series, we perform an augmented Dicky-Fuller test and investigate the serial autocorrelation using a Ljung-Box (LB) test with lag one (eight hours) and three (one day) for each parameter time series. Our ADF test results indicate that the unit root hypothesis can be rejected for all parameters at the highest significance level, which indicates that each beta time series is stationary.

The mean and standard deviation of each beta time series for both models depict similar, albeit slightly different absolute values. For instance, the mean of β_0 is identical for both at -0.344 which points towards a general surface level of 70.89%. This value is in close proximity to the 10-days ATM IV average over the sample period of 71.57%, as previously noted. Indeed, correlation between Model 1 (Model 2) β_0 and the constant 10-days ATM implied volatility is 0.94 (0.94). Additionally, the standard deviation for both models is nearly identical at a relatively high level, reflecting wide fluctuations in the base surface level caused by abrupt (downward) movements in the underlying asset, such as the COVID crash in March 2020.

Both models depict a negative mean for β_1, implying a negative relationship between strike level and implied volatility with a negative β_1 in 79.92% (82.56%) of the time. Combining this with the positive value for β_2 demonstrates a declining implied volatility in moneyness but increasing in squared moneyness. This indicates an asymmetric volatility smirk, pronounced more for OTM put than calls as seen in Figure 7.1 and consistent with the typical equity index implied volatility shapes (Neumann and Skiadopoulos, 2013). Similar, the positive mean for β_3 illustrates an upward movement for increasing maturity, observed in 78% (79%) of the time. In contrast, the β_4 mean is negative, as discussed earlier, and has high correlation with β_3 (-0.95 for both models). A linear increasing β_3 requires a concave, i.e. negative β_4, to approach the yield curve pattern of volatility term structure.

---

[9]For instance, considering Model 1 (Model 2) we see an improvement of the average RMSE from 0.047 to 0.039 (0.045 to 0.037); the R^2 value increases from 0.886 to 0894 (0.898 to 0.906), the adjusted R^2 enhances from 0.88 to 0.891 (0.886 to 0.902) as well as AIC from -372 to -502 (-382 to -515) and BIC from -359 to -488 (-367 to -497).

Figure 7.3: Model 1 Time Series β Coefficients

Time series of each β coefficient of the deterministic implied volatility surface model estimated using the robust ordinary least square method on Model 1, covering our sample period from August 2019 to August 2022 in eight hour steps. We omit the horizontal time axis for the sake of clearly but keep it for the last panel. The vertical axis is unit-free and changes for each parameter.

Figure 7.4: Model 2 Time Series β Coefficients

Time series of each β coefficient of the deterministic implied volatility surface model estimated using the robust ordinary least square method on Model 2, covering our sample period from August 2019 to August 2022 in eight hour steps. We omit the horizontal time axis for the sake of clearly but keep it for the last panel. The vertical axis is unit-free and changes for each parameter.
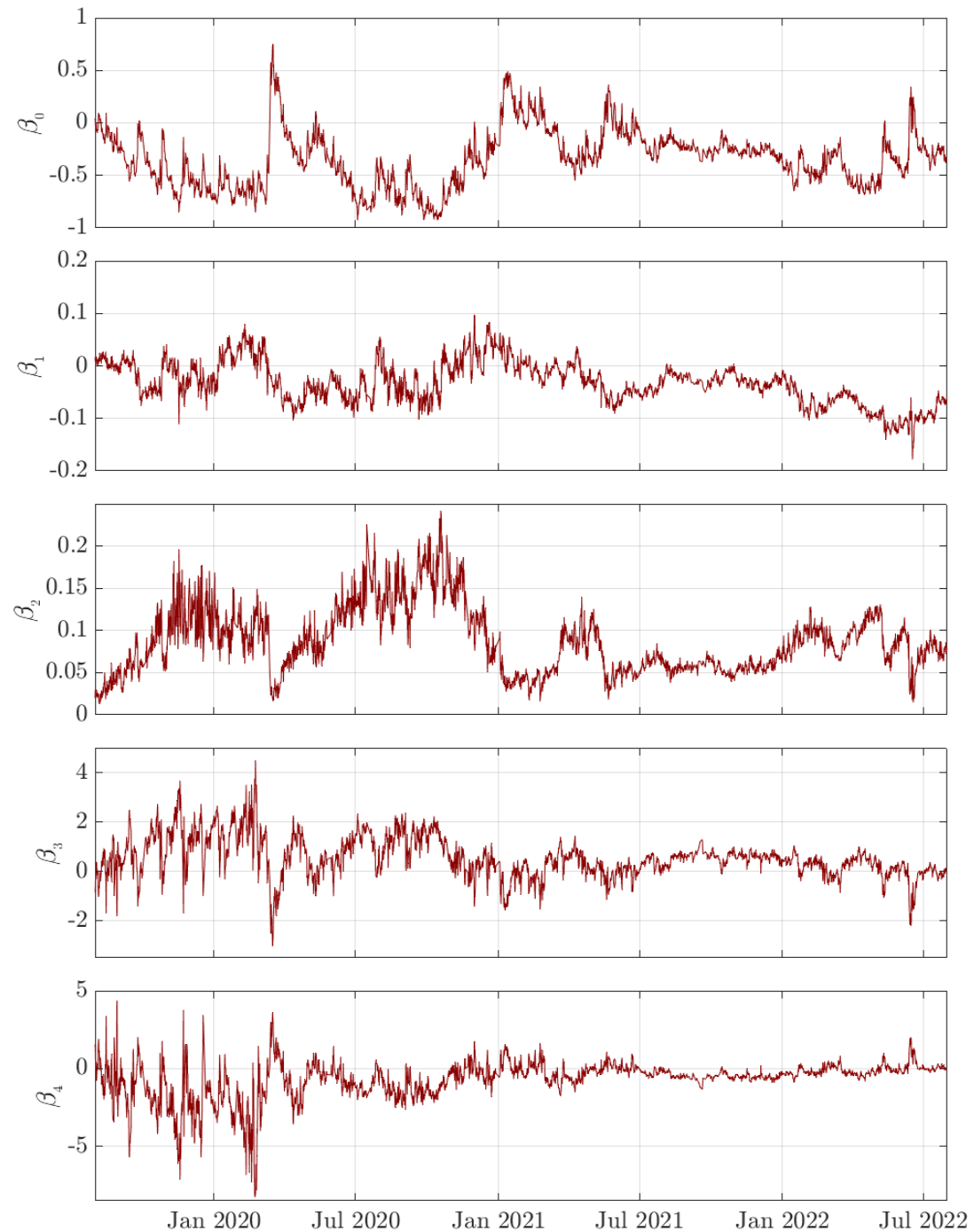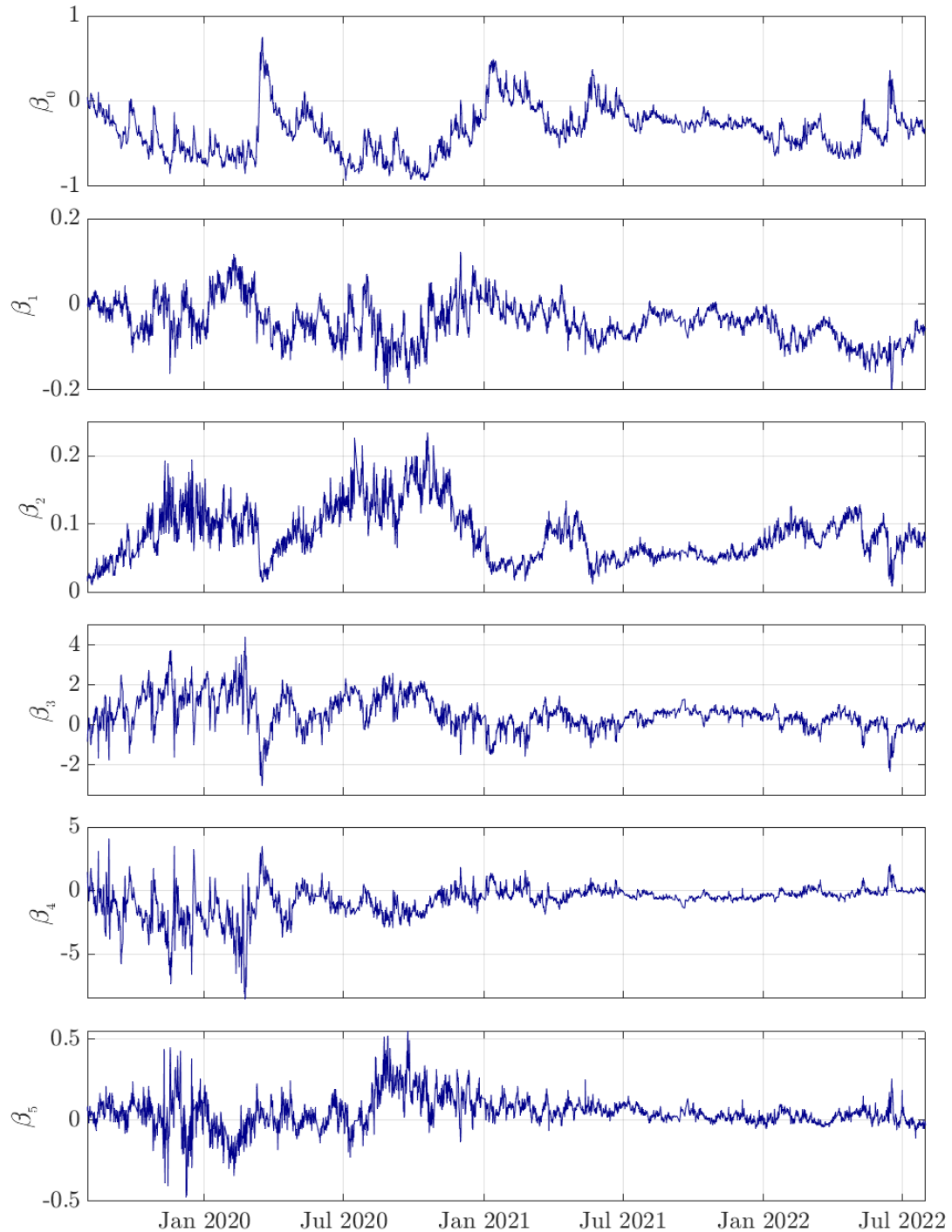
Table 7.2: Coefficient Time Series Properties

Summary statistics of the deterministic surface coefficients. It depicts the mean and standard deviation of each beta coefficient time series estimated by Model 1 (7.6; upper panel) and Model 2 (7.7; lower panel) via robust OLS. We further depict the average value of a two-sided t-Test statistic for each parameter across all regressions. The percentage of significant value with a standard 10% size for each parameter is presented in parentheses underneath. Furthermore, we present the ADF test value to investigate for stationarity and a Ljung-Box (LB) test with lag one and three to asses for autocorrelation of the coefficient time series. We illustrate statistical significance with the asterisk at the 1% *** and 5% ** level. The presented results cover the whole sample data, i.e. three years of eight-hour options data, starting in August 2019

Coefficient	Mean	Std. Dev.	t-Test	ADF	LB(1)	LB(3)
Model 1						
β_0	-0.344	0.269	-25.59 (98.2)	-4.248^{***}	2978^{***}	8602^{***}
β_1	-0.033	0.04	-5.51 (82.46)	-3.004^{**}	2865^{***}	8236^{***}
β_2	0.087	0.039	13.37 (99.97)	-3.55^{***}	2752^{***}	7896^{***}
β_3	0.539	0.813	4.01 (86.54)	-5.061^{***}	2659^{***}	7199^{***}
β_4	-0.694	1.222	-3.05 (82.04)	-5.263^{***}	2513^{***}	6589^{***}
Model 2						
β_0	-0.344	0.27	-23.38 (98.46)	-4.243^{***}	2978^{***}	8603^{***}
β_1	-0.043	0.048	-4.93 (82.46)	-3.751^{***}	2673^{***}	7456^{***}
β_2	0.085	0.0381	13.2 (99.90)	-3.592^{***}	2727^{***}	7805^{***}
β_3	0.56	0.821	4.35 (88.18)	-4.96^{***}	2671^{***}	7246^{***}
β_4	-0.738	1.219	-3.47 (85.22)	-5.051^{***}	2517^{***}	6605^{***}
β_5	0.048	0.098	1.58 (56.50)	-5.952^{***}	2119^{***}	5863^{***}

As stated previously, the main objective of this chapter is to explore the predictability of coefficients driving the implied volatility surface. To verify the presence of serial autocorrelation, we conduct a Ljung-Box (LB) test with lag one (eight hours) and three (one day) for each parameter time series. Our results in Table 7.2 indicate that each parameter is highly autocorrelated at the highest significance level, regardless of the model. These findings support and motivate the use of Vector Auto-Regressive (VAR) models as a suitable tool for analysing and predicting the dynamics and interrelationships among the parameters.

7.2.2 Predictive Models

The core of our prediction analysis lies in modelling the $\boldsymbol{\beta}$ time series. For this, we compare three distinct approaches – naive, statistical, and machine learning – and select the best model on its mean squared error (MSE) and R^2 value. In the following, we keep the notation for the observed beta vector at time t as $\boldsymbol{\beta_t}$ and denote any prediction or estimate as $\hat{\boldsymbol{\beta}}_t$. To evaluate model performance and account for real-world conditions, we divide our data into three sets: an in-sample (or train) set, a validation set, and an out-of-sample (or test) set. This division is based on a 30:6:5 month ratio and covers the periods August 2019 to January 2022, February 2022 to July 2022, and August 2022 to end December 2022, respectively. We adopt this ratio in line with the recommendation of Hastie et al. (2009), which allows for robust training of potential relationships while accommodating different regimes. For each of the models, we define and calibrate the optimal model parameter using the first 73% of the data, validate the optimal hyperparameters using the subsequent 15%, and apply the model to the remaining 12%. When appropriate, we will update the models recursively following each new time step.

Naive Model

Most academic papers on predictive model forecasting fail to provide a baseline which is both simple and does not require much computational power. Without a baseline, sophisticated models may be set up needlessly, as a 'quick and dirty' or *naive* approach might achieve same or better results. A naive forecasting model is a very simple and deterministic framework which requires only a minimum of input and set-up costs. If any more sophisticated model fails to exceed this baseline, the approach should be abandoned. To this end, we follow both Gonçalves and Guidolin (2006) and Bernales and Guidolin (2014) and apply the 'ad-hoc strawman' model. This model goes back to Christoffersen (2004) and is simply a Random Walk (RW) process for each beta time series. Under this model, the best prediction for the next time step is the current value with a noise term:

$$\hat{\boldsymbol{\beta}}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\varepsilon}_t,$$

where $\boldsymbol{\varepsilon}_t$ is a normal $\Phi\left(0, \Sigma_\beta\right)$-distributed noise-term. The risk horizon for variance is a crucial element. To obtain an accurate estimation, we define Σ_β as a vector of rolling window realised variances for each beta parameter return over the past p lags.

We compare different lags p, corresponding from day up to two months, respectively. A longer history may introduce spurious – or ghost – features, particularly for the rolling window approach, which could bias the results. We use the `math` library to evaluate the MSE and R^2 and calculate the realised variance using the `random.normal` function of `numpy` to simulate the normal-distributed error. Note that simulations are generated randomly and no two are identical, hence, to enhance the reliability of our results, we adopt the average of

five simulations. We test the random walk model with different lags, ranging from one day to three months, on the validation set and find that using the last month returns' realised variance outperforms any other lag in terms of MSE.[10].Therefore, we set the window size to 30 days and consider the random-walk model to be the baseline.

Statistical Model

There exists a plethora of statistical techniques to model any type of time series. However, given the specific problem we face, i.e. a multivariate stationary time series with serial auto-correlation, we employ a Vector Auto-Regressive (VAR) model. The rationale behind this is, due to autocorrelation in time series, the relevant information for making a prediction is often within a small window of past observations. The simplistic nature of linear models make them a prominent class of time series prediction models. Granger and Newbold (1974) are among the first to introduce the Vector Auto-Regressive model of lag p which is a generalisation of the univariate AR model (Box and Gwilym, 1970) allowing multiple time series to intact with each other linearly. In this setup, each variable is modelled as a linear function of its own past p lags as well as the past values of other variables in the system. For this, we consider the time series of the beta vector $\boldsymbol{\beta}$ and forecast the step-ahead coefficients by:

$$\boldsymbol{\beta_t} = \boldsymbol{\gamma} + \sum_{i=1}^{p} \boldsymbol{\Psi}_i \boldsymbol{\beta_{t-i}} + \boldsymbol{u_t},$$

where $\boldsymbol{\gamma}$ is the intercept vector, $\boldsymbol{\Psi}_i$ are the coefficient matrices capturing the linear relationship between the responsive variable and its own p lags, and $\boldsymbol{u_t} \sim \Phi\left(0, \boldsymbol{\Sigma_t}\right)$ is the i.i.d. distributed error term. Here, Φ denotes the normal distribution where $\boldsymbol{\Sigma_t}$ is the covariance matrix of the individual error terms. The assumption of this VAR(p) model, which follows a multivariate normal distribution with a zero mean and constant covariance matrix, may not hold over time as the covariance matrix $\boldsymbol{\Sigma_t}$ changes. In order to address this, we employ a recursive daily refitting strategy, in which we use p lags to predict $\hat{\boldsymbol{\beta}}_{t+1}$, then observe the real value and use again the last p lags at $t+1$, i.e. $\boldsymbol{\beta_{t+1-p}}, ..., \boldsymbol{\beta_{t+1}}$ to predict $\hat{\boldsymbol{\beta}}_{t+2}$, and so on, for three observations. By continuously refitting the model daily this way, the assumption of constant $\boldsymbol{\Sigma_t}$ is more reasonable and the resulting predictions more accurate. Note that we assure the stationarity of the data as we include more days and integrate the time series if necessary. In order to ensure a fair comparison between models, we select p based on the best performance on the validation set. Our analysis covers the same lag size as before, i.e. $p = 1, ..., 180$, and indicates that lag $p = 75$ (79) yields the minimal MSE of 0.0017 (0.0016) for Model 1 (2), respectively. For the technical implementation, we use the time series analysis library (`tsa`) of `statsmodel` to fit the VAR model with different lags.

---

[10]The random walk model with $p = 90$ lags achieves an average MSE of 0.0108 and $R^2 = 82.69\%$, outper-forming the other kags

Machine Learning Model

The machine learning model completes our group of predictive models. Here again, there exists a plethora of different models under the umbrella of supervised, unsupervised or reinforcement learning. First, we need to understand that we face a regression problem and we want to model the time series in order to obtain predictions. Unsupervised learning models are not suited for this task as they are generally used for clustering or association problems. By the same token, reinforcement learning models are not ideal as these aim to solve Markovian decision problems, i.e. it provides an optimal behaviour for an agent in an environment with different states and actions. As such, it does not model or predict the time series. The supervised learning category can be split up into a classification and regression subclass. Particularity the latter is of interest for us. Here, a wide cohort of scholars and practitioners (D'Amato et al., 2022; Brownlee, 2021) advocate the Long-Short Term Memory model.[11] Given the success of the LSTM application to implied volatility data (Medvedev and Wang, 2022; Bolch and Book, 2022), we decide to add a LSTM model as final predictive model to our portfolio.

The LSTM model is a special type of recurrent neural network which can process and predict sequential data effectively while accounting for historical dependencies between the input variable, see Chapter B of the Appendix for an in-depth description of the model architecture. As such, it promises to learn complex non-linear relationships between past lags and beta parameter by defining a mapping function which takes historical beta as input and yields a prediction. Many scholars see it as a natural non-linear extension of VAR-type models. In this framework, the vector forecast is given by:

$$\hat{\boldsymbol{\beta}}_t = f\left(\boldsymbol{\omega}, \boldsymbol{\mathcal{B}}_{t-1}\right),$$

where f denotes the mapping function of the machine learning algorithm for a given internal, fully trained set of weights $\boldsymbol{\omega}$, and $\boldsymbol{\mathcal{B}}_{t-1}$ describes the (#samples, #lags, #features)-dimensional input tensor, i.e we provide a $(1, p, 5)$-dimensional input Model 1, and $(1, p, 6)$ for Model 2, respectively. In light of the regression problem in hand, we choose the MSE as designated loss function and use the adaptive moment estimation (`adam`) optimisation algorithm (Kingma and Ba, 2015) for the optimisation process. This optimisation algorithm is widely accepted and used in the industry for its efficient use of memory and speed. Other loss functions such as the log or hinge loss are more suitable for classification problems. An alternative could have been the mean absolute loss, but given the broad acceptance by practitioners, we decide to proceed with the industry standard for regression problems.

Initially, we scale each time series to [0,1] individually, and use the same scaling factor to invert the prediction to interpretative values. During the training process, it might be that

---

[11]There exist different variations of the plain-vanilla LSTM model such as the convolutional LSTM or Transformer (Vaswani et al., 2017) which have shown great results in financial time series and natural language processing. However, these models require an amount of data to train which simply cannot provide with our database.

the best fit, i.e. the optimal model weights, is found not at the end, but prior. To obtain the best possible overall model and not waste time/computational energy after the model converged, we add `CheckPoint` and `EarlyStopping` to our algorithm. In our case, we observe the loss function and stop the training iff there is no improvement over the span of 20 epochs. We determine the optimal set of hyperparameters, e.g. 103 input samples, 200 epochs, 150 nodes, and a batch size of 64 based on our grid-search approach across over 1200 different permutations. Furthermore, we apply the Monte Carlo Dropout (MCDO) and compute each time step in our prediction based on the average of ten simulations, taking into account a trade-off between computational efficiency and accuracy. Our model achieves a R^2 value of 88.38% (80.23%) for the train (validation) set and a MSE of 0.0019 (0.0014) and refer to Section C.1 of the Appendix for a detailed description of the choice of parameter.

The prediction procedure is similar to the VAR case, i.e. it involves training the model with all available data, followed by generating predictions for the next time step ($\hat{\boldsymbol{\beta}}_{t+1}$) and comparing these to observed values ($\boldsymbol{\beta}_{t+1}$). Subsequently, we use the observed values as new input to generate additional predictions ($\hat{\boldsymbol{\beta}}_{t+2}$). Model retraining occurs only after three consecutive observations, corresponding to a one-day interval, and this process is repeated daily over the period from August to December 2022. Again, we ensure that the updated data is stationary and update the scaling factors. Similar to the previous models, all coding is done in Python using the `TensotFlow` and `Keras` libraries.

7.2.3 Performance Measures

Both statistical and economic measures play a critical role to evaluate the performance of our forecasting models. We use statistical measures to assess the accuracy and reliability of our model's forecasts, while the economic side explain the profitability of the forecasts. Our statistical measures include root mean square error, mean absolute error, among others others; and the profitability of three distinct trading strategies based on the forecasting signal is measured on PnL and return-ratios. First, we translate the beta predictions using formulae (7.6) and (7.7) into implied volatilities and replicate the predicted surface. For the statistical analysis of the model performances, we follow Bernales and Guidolin (2014) and define the following measures:

- **Root Mean Squared Error (RMSE):** This is the average of the square root of the squared deviations of market implied volatilities from the model's forecast implied volatilities, averaged over the number of options traded. This gives insight about the distance between forecast and observed IV.

- **Mean Absolute Error (MAE):** This is the prediction error in implied volatilities as the average of the absolute differences between the market IV and the model's forecast implied volatility across traded options. This considers the correct direction of the prediction as well as the distance.

- **Mean Correct Direction (MCD):** This is the average frequency (percentage of observations) for which the change in implied volatility predicted by the model has the same sign as the realised change in implied volatility indicates.

- **Diebold and Mariano (1995) (DM)**: This test statistic measures the forecasting performance and significant differences of different models. It tests the null hypothesis that two models have the same forecast accuracy. If the DM test statistic exceeds the critical value, we can reject the null hypothesis and conclude that one model outperforms the other with statistical significance.

The second part of our study aims to evaluate the effectiveness of our forecasts in supporting portfolio decisions. We introduce a signal $q_t(m, T)$, which represents the difference between the expected or forecasted implied volatility and the current market implied volatility at an arbitrary grid point (m, T) on the implied volatility surface at time t. The signal is positive (negative) if the model forecasts an increase (decrease) of the implied volatility at (m, T) from t to $t + 1$. As the implied volatility reflects the price of an option, we will buy (sell) this option when receiving a positive (negative) signal. We denote the set of options receiving a positive signal, i.e. a buy sign, as:

$$Q_t^+ = \{(m, T) : q_t(m, T) > 0\},$$

as well as the options receiving a negative signal

$$Q_t^- = \{(m, T) : q_t(m, T) < 0, \}.$$

Note that we consider the implied volatilities with over eight decimal numbers. Hence, in no occasion we find that that the predicted implied volatility matched the current IV.

We propose three distinct trading strategies to assess the practicality of the signals, assuming perfect fractional investing is possible, and that we can enter and exit a position at each time step. The strategies involve ATM straddles, delta-hedged portfolios and a baseline model. For each strategy, we assume that \$1000 are invested initially, and evaluate the resulting profit and loss of these hypothetical strategies, comparing them with each other based on their mean return, standard deviation as well as Sharpe- and Sortino ratio. Note that the forecasts will never *exactly* match the implied volatility, i.e. the forecast will always provide an indication of direction and magnitude of the volatility. Hence, our focus will be on strategies involving directional position of future volatility and those addressing the predicted magnitude of the IV. We follow Harvey and Whaley (1992) with our strategies who also highlight that these strategies are an ideal way to reduce the transaction costs.

Strategy A

The first trading strategy (Strategy A) involves buying or selling ATM straddles with varying maturities. This approach excludes considering strike levels away from ATM, as ITM options are rarely traded for both call and put options. An ATM straddle involves purchasing (or selling) a call and put option with the same strike price and maturity, creating an instantaneous delta-neutral position that is a pure bet on future volatility. In the event of a negative signal at a given maturity, we sell an ATM call and put option, thereby placing a bet on a decrease in future volatility. Conversely, for a positive signal, we take a long position in the ATM straddle since we expect an increase in future volatility. Note that we focus solely on the directional forecast of the signal, not its magnitude. Therefore, unless the models predict different directions, the profit and loss will be identical across all models. Since the moneyness of an option is rarely exactly 1, we select the closest moneyness to $m = 1$ that has both calls and puts traded. We then consider special subsets $\widetilde{Q_t}^+$ and $\widetilde{Q_t}^-$ for this trading strategy:

$$\widetilde{Q_t}^{(+,T)} = \left\{ (m,T) \in Q_t^+ : m = \underset{\hat{m}}{\arg\min} |1 - \hat{m}| \right\},$$
$$\widetilde{Q_t}^+ = \bigcup_{T \in \mathcal{T}} \widetilde{Q_t}^{(+,T)}$$

and

$$\widetilde{Q_t}^{(-,T)} = \left\{ (m,T) \in Q_t^- : m = \underset{\hat{m}}{\arg\min} |1 - \hat{m}| \right\},$$
$$\widetilde{Q_t}^- = \bigcup_{T \in \mathcal{T}} \widetilde{Q_t}^{(+,T)},$$

where $\mathcal{T}$ describes the set of all tradable maturities. The straddle strategy is straight forward. Let V_t^A be the portfolio value of applicable straddles:

$$V_t^A = \sum_{x \in \widetilde{Q_t}^+} [C_t(x) + P_t(x)] - \sum_{x \in \widetilde{Q_t}^-} [C_t(x) + P_t(x)],$$

where $C_t(m,T)$ and $P_t(m,T)$ are the observed market prices of call and put option with moneyness m and maturity T, respectively. For a positive (negative) portfolio value we purchase (sell) X_t^A units of the straddle portfolio:[12]

$$X_t^A = \frac{\$1000}{|V_t^A|}$$

---

[12]Note that we assume zero interest rate. Otherwise, we could invest the initial capital into the risk free rate until the next time step in case of $V_t^A < 0$ where we would sell the quantity X. Given the high rebalancing frequency of this study and the current interest rate, the risk-free profit generated within such a short window is negligible and does not change the results of the study.

and calculate the dollar-value of the profit G as:

$$G_{t+1}^A = X_t^A \left[\sum_{x \in \widetilde{Q}_t^+} \left(\left[C_{t+1}(x) + P_{t+1}(x) \right] - \left[C_t(x) + P_t(x) \right] \right) \right]$$
$$+ X_t^A \left[\sum_{x \in \widetilde{Q}_t^-} \left(\left[C_t(x) + P_t(x) \right] - \left[C_{t+1}(x) + P_{t+1}(x) \right] \right) \right],$$

and subsequently the percentage return as:

$$\mathcal{R}_{t+1}^A = \frac{G_{t+1}^A}{1000}.$$

Strategy B

The second strategy (Strategy B) involves a straightforward delta-hedged position responding to each signal. For this strategy, we consider the entire options chain for each tradable maturity and purchase/sell each call or put option based on the corresponding signal q. Moreover, we establish an opposing position in the underlying S_t in the size of the option's delta. To achieve this, we define novel subsets for the call options that we purchase $\left(\widetilde{Q}_t^{+,C} \right)$, the calls we sell $\left(\widetilde{Q}_t^{-,C} \right)$, and puts we buy $\left(\widetilde{Q}_t^{+,P} \right)$ and puts we sell $\left(\widetilde{Q}_t^{-,P} \right)$, respectively as:

$$\widetilde{Q}_t^{+,C} = Q_t^+ \setminus \{(m,T) : m < 1\}, \qquad \widetilde{Q}_t^{+,P} = Q_t^+ \setminus \{(m,T) : m \geq 1\}$$
$$\widetilde{Q}_t^{-,C} = Q_t^- \setminus \{(m,T) : m < 1\}, \qquad \widetilde{Q}_t^{-,P} = Q_t^- \setminus \{(m,T) : m \geq 1\}.$$

The value of the delta-hedged portfolio is given by:

$$V_t^B = \sum_{x \in \widetilde{Q}_t^{+,C}} \left[C_t(x) - S_t \delta_t^C \left(x, \sigma_{t+1}^\xi \right) \right] - \sum_{x \in \widetilde{Q}_t^{-,C}} \left[C_t(x) - S_t \delta_t^C \left(x, \sigma_{t+1}^\xi \right) \right]$$
$$+ \sum_{x \in \widetilde{Q}_t^{+,P}} \left[P_t(x) + S_t \delta_t^P \left(x, \sigma_{t+1}^\xi \right) \right] - \sum_{x \in \widetilde{Q}_t^{-,P}} \left[P_t(x) + S_t \delta_t^P \left(x, \sigma_{t+1}^\xi \right) \right],$$

where $\delta_t \left(m, \tau, \sigma_{t+1}^\xi \right)$ is the option's delta with strike m and TTM τ, calculated using each models' step-ahead predicted implied volatility σ_{t+1}^ξ with $\xi \in$[RW, VAR, LSTM].[13] Here again, we purchase (sell) the quantity X_t^B in units of the delta-hedge portfolio V_t^B. The same course of action applies for the dollar-profit:

---

[13]We make an additional assumption following both Gonçalves and Guidolin (2006) and Bernales and Guidolin (2014) who argue that it is a common practice for market maker to assume the same value for the underlying at $t+1$ as in t. This is also in line with the efficient market hypothesis.

$$G^B_{t+1} = X^B_t \left[\sum_{x \in \widetilde{Q_t}^{+,C}} \left(\left[C_{t+1}(x) - S_{t+1}\delta^C_t \left(x, \sigma^\xi_{t+1} \right) \right] - \left[C_t(x) - S_t\delta^C_t \left(x, \sigma^\xi_{t+1} \right) \right] \right) \right]$$

$$+ X^B_t \left[\sum_{x \in \widetilde{Q_t}^{+,P}} \left(\left[P_{t+1}(x) - S_{t+1}\delta^P_t \left(x, \sigma^\xi_{t+1} \right) \right] - \left[P_t(x) - S_t\delta^P_t \left(x, \sigma^\xi_{t+1} \right) \right] \right) \right]$$

$$+ X^B_t \left[\sum_{x \in \widetilde{Q_t}^{-,C}} \left(\left[C_t(x) - S_t\delta^C_t \left(x, \sigma^\xi_{t+1} \right) \right] - \left[C_{t+1}(x) - S_{t+1}\delta^C_t \left(x, \sigma^\xi_{t+1} \right) \right] \right) \right]$$

$$+ X^B_t \left[\sum_{x \in \widetilde{Q_t}^{-,P}} \left(\left[P_t(x) - S_t\delta^P_t \left(x, \sigma^\xi_{t+1} \right) \right] - \left[P_{t+1}(x) - S_{t+1}\delta^P_t \left(x, \sigma^\xi_{t+1} \right) \right] \right) \right],$$

and similar to the first strategy, we calculate the returns as:

$$\mathcal{R}^B_{t+1} = \frac{G^B_{t+1}}{1000}.$$

Strategy C

In addition to these two strategies, we want to include a passive strategy. We consider the infamous buy and 'hodl' strategy. This term has become popular among the crypto community and describes a simply buy and hold strategy, sometimes referred to as *hold on for dear life*. This will be the benchmark the other two strategies will have to outperform. The PnL of this strategy is given by:

$$\mathcal{R}^C_{t+1} = 1000 \times \left(\frac{S_{t+1}}{S_t} - 1 \right).$$

We want to measure the performance of the three strategy first and foremost based on their return and standard deviation. In addition, we consider the standard deviation, Sharpe (1994)- and Sortino and Price (1994)-Ratio of the strategies. Particularly the latter two are important performance measures for any trading strategy as they provide a quantitative assessment of the return generated by our trading strategy relative to the risk taken to achieve that return.[14] Traders are interested in these ratios as they provide a way to compare the performance of different strategies or investment opportunities on a risk-adjusted basis. A high Sharpe or Sortino ratio indicates that the returns are high relative to the level of risk taken, while a low ratio indicates the opposite. A low or negative Sharpe or Sortino ratio suggests that the returns are not worth the risk taken, and we should abandon the strategy.

---

[14]The Sharpe ratio measures the excess return generated by the strategy per unit of risk taken, where risk is measured as the standard deviation of returns. The Sortino ratio, on the other hand, is a variation of the Sharpe ratio and considers only downside risk, which is the risk of losses, i.e. it only accounts for downside volatility.

7.3 Empirical Results

Motivated by the discussions in Section 7.2, we present the empirical results in this section. Initially, we evaluate the out-of-sample prediction efficiency of the random walk, VAR and LSTM model using the measures introduced earlier; and subsequently, we investigate the profitability of these forecast signals through the three trading strategies.

7.3.1 Statistical Results

Figure 7.5 presents the step-ahead predictions of each beta coefficient from the random walk (purple), VAR (orange) and LSTM (green) models, in addition to the actual observed Model 1 beta values (red) during the out-of-sample period August to December 2022. In general, all models capture trends effectively and generate reasonably precise predictions for all five parameters. The figure shows that the random walk model outperforms the two more sophisticated models during stable tranquil (August to November) but exhibits substantial fluctuations following significant jumps in the time series, mainly due to the ghost-features of the rolling window. Particularly for the latter two parameter, which depict a generally wider range, the RW demonstrates wide deviations from the true value around December 2022. The VAR model captures the true value accurately but appears to lag slightly during larger movements as it struggles to grasp substantial jumps and requires 2-3 time steps to adjust. In contrast, the LSTM model yields similar good forecasts and is less prone to wide deviation during sudden jumps. In fact, on rare occasions, such as 8 November or 16 December, it even predicts the jump in the underlying bitcoin price successfully. This could be due to the internal memory of the model, i.e. it might capture the current regime better, remember it from previous observations, and use it better than the other two and react accordingly.

The statistical analysis of the forecasted values with the observed beta time series leads to the conclusion that the VAR and LSTM outperform the random walk model throughout all beta parameters in terms of RMSE and MAE.[15] The main contributor to the poor performance of the RW is certainly the inevitable ghost-features of a rolling window. Both the VAR and LSTM models are better equipped to handle changes. However, we cannot identify one definitive superior model between the VAR and LSTM models. In fact, the LSTM approach captures the movements of β_0, β_3 and β_4, i.e. the general surface level and term structure more accurately than the VAR. In contrast, the VAR model outperforms for the volatility skew parameter β_1 and β_2 which supports the findings of Lewis (2000). The Model 2 findings of the analysis are consistent for the out-of-sample predictions, indicating robustness of the results. The results of the six factor model are presented Section C.2.

---

[15]The mean absolute error for the random walk model for each beta coefficient, starting with β_0 to β_4 is 0.042, 0.006, 0.007, 0.198 and 0.153. By the same token, the mean absolute error of the VAR model are 0.0189, 0.003, 0.002, 0.044, 0.082 and the LSTM model shows 0.014, 0.005, 0.004, 0.041, 0.05. Similar, for the RMSE error we identify for the RW 0.058, 0.009, 0.01, 0.281, 0.287; for the VAR 0.054, 0.009, 0.01, 0.2, 0.216; and LSTM 0.031, 0.009, 0.01, 0.005, 0.075.

Figure 7.5: Out-of-Sample Time Series $\boldsymbol{\beta}$ Coefficients

Time series of each $\boldsymbol{\beta}$ coefficient (red) of the deterministic implied volatility surface model estimated using the robust ordinary least square method on Model 1, covering our out-of-sample period from August 2022 to December 2023 in eight hour steps as well as the step-ahead prediction using the random walk benchmark (purple), Vector Auto-Regressive model (orange) and Long-Short-Term memory model (green). We omit the horizontal time axis for the sake of clearly but keep it for the last panel. The vertical axis is unit-free and changes for each parameter.
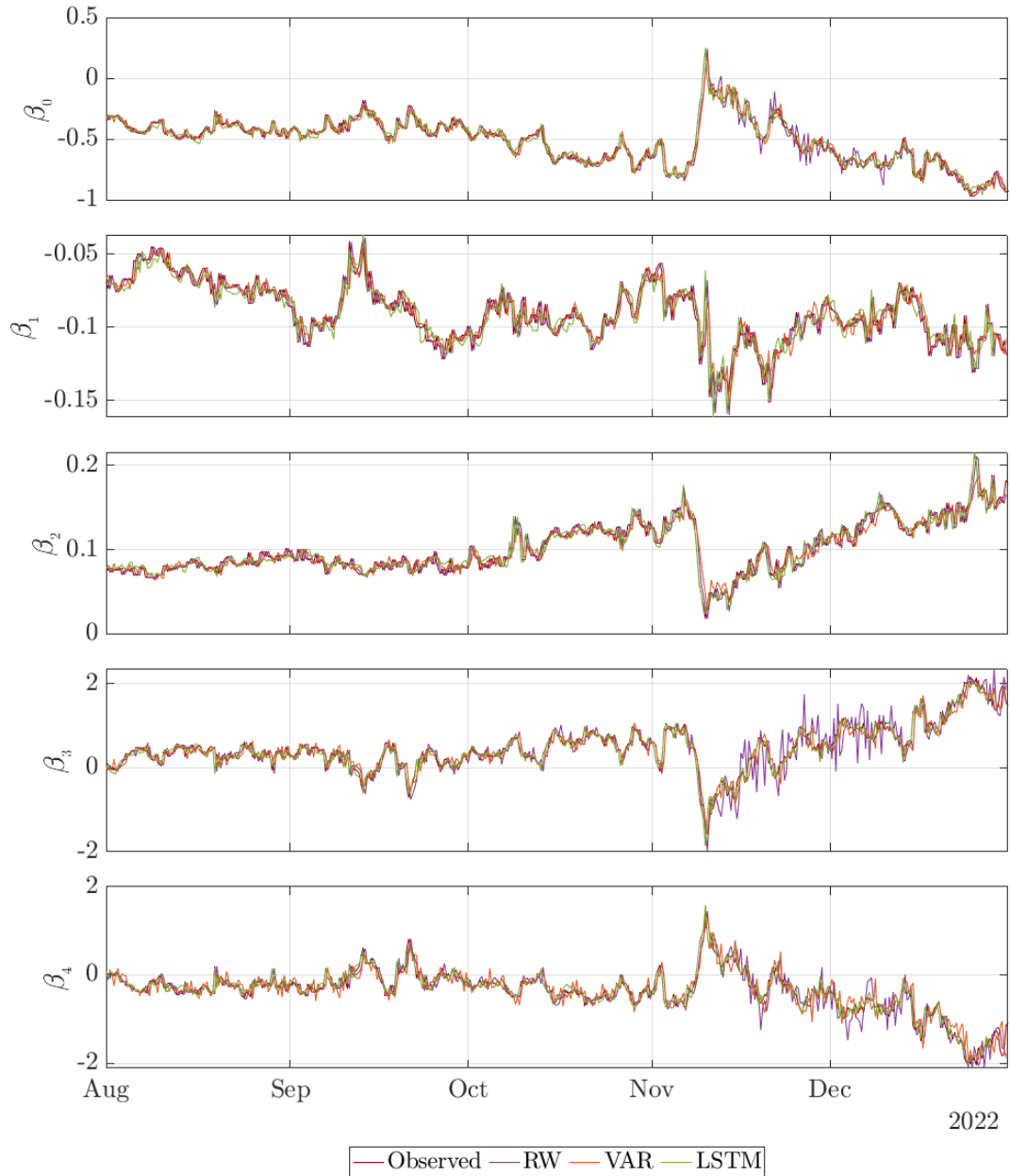
Table 7.3 presents the prediction errors of implied volatilities derived from the beta fore-casts. Specifically, we translate the forecasted beta vector into individual option implied volatilities at each time step and compare them with the actual observed implied volatility in the subsequent time step. Our performance evaluation measures the predictive accuracy using root-mean-squared error, mean absolute error, and mean directional error, and we report the results in both absolute values (RMSE, MAE) and percentage points (MCD) for all step-ahead predictions throughout the out-of-sample period. To explore potential variations across strike levels and maturities, we further partition the predictions into distinct subgroups, including short-, mid-, and long-term maturities, as well as OTM call/put and ATM options. Our choice of the moneyness- and maturity range is motivated by the trading volume discussed in Section 3.2. The exact maturity ranges are identical to Chapter 3, i.e. up to two weeks, between two weeks a month and more than a month.

Our findings present a nuanced picture. Considering the entire surface, the VAR model exhibits the lowest RMSE and MAE among the tested models but delivers an unsatisfying MCD error. This observation remains consistent across different moneyness ranges. In fact, the VAR model surpasses the LSTM model by 11.3% (10.6%) with respect to the RMSE (MAE), and outperforms the RW even by 17.7% (21.8%). Both RW and VAR demonstrate rather arbitrary prediction patterns, accurately forecasting the direction in 50.8% and 51.8% of the time, respectively. Only the LSTM distinguishes itself in this aspect with 55.3%. It is not surprising that the RW approach demonstrates the weakest performance overall and for each moneyness/maturity subcategory in each metric. After all, a random prediction does not take into account momentum and the ghost features' influence on the overall results are severe. Table 7.3 highlights the findings of Figure 7.5, i.e. VAR models the smirk better while the LSTM captures the term structure more accurate.

We find remarkable results considering ATM strike levels across all maturities. The VAR model's long-term ATM prediction yields the overall best RMSE (0.025) and MAE (0.018) values, however, the directional accuracy appears rather random. In contrast, the LSTM shows great accuracy, despite having marginally higher RMSE and MAE. It is evident that, as maturity decreases, the LSTM model's accuracy in predicting the correct direction improves. Specifically, while the model predicts the direction of long-term maturities in 52.6% of cases accurately, this increases to 59.6% for mid-term options and peaks at 64.6% for short-term ATM options. It is only in this subgroup that the VAR model produces higher RMSE and MSE (both +11%) than the LSTM, while its directional accuracy remains rather arbitrarily at 53.9%. Note that the correct direction is of paramount importance.[16]

Across all maturity classes, we observe considerable discrepancies between the tail predic-tions for both call and put options and ATM implied volatilities. Irrespective of maturity and model, tails exhibit higher RMSE and MAE values. Similarly, the directional forecast appears

---

[16]Consider this simple example: A stock price is currently traded at $100. Model A predicts the price to go down to $99 while Model B forecasts $110. The stock price moves to $101; Model A has lower RMSE and MAE but any trader would have lost money relying on the first prediction.

to deteriorate as we progress further OTM, with the sole exception of mid- and long-term OTM puts. The most plausible explanation for this phenomenon is the insufficiency of the second-order Taylor expansion to capture the tail adequately. This issue is, in fact, twofold: First, bitcoin option trading is predominantly clustered around ATM and near-maturity options, as demonstrated in Section 3.2. It is crucial to understand that, in an OLS, each data point is assigned equal weight. Consequently, the concentration of observations around short-term ATM options may lead to inaccurate model fits for OTM or longer-dated options. Second, the tails exhibit greater shape variation, particularly for OTM calls. While the OTM put implied volatility keeps a decreasing trend for increasing moneyness, the OTM call IV frequently transitions between a (symmetric) smile and a hockey-shaped smirk. This allows for more consistent modeling of OTM put implied volatility, while capturing OTM call implied volatility proves to be more challenging.

Table 7.4 presents the outcomes of the Diebold-Mariano test statistic which conducts a pairwise comparison of the forecast accuracy of the three models, i.e. RW-VAR, RW-LSTM and VAR-LSTM. A positive DM value implies a preference for the second model over the first. For example, the DM test statistic of 7.4 for the VAR-LSTM pair for mid-term ATM options indicates that the LSTM model outperforms the VAR model in predictive performance. Our findings here confirm the statistical evaluation from Table 7.1, with most results exhibiting the highest significance. In other words, no strike/maturity subgroup demonstrates a preference for the RW model, and the VAR model outperforms its counterparts overall. Nevertheless, specific maturities and moneyness levels, particularly within the short- and mid-term maturity ranges, yield different outcomes, such as the LSTM model outperforming the other two models. We find the most pronounced results for short-term ATM options, which exhibit a strong preference for the LSTM model, albeit to a lesser degree on the tails. The findings of Tables 7.1 and 7.4 motivate us to consider short-term ATM options for our further empirical research. This particular subclass exhibits minimal errors while demonstrating the highest proportion of accurate directional forecasts. Pursuing any other subclass for potential trading strategies would likely generate inferior results, hence make them unworthy of further consideration.

Particularly the LSTM model seems to outperform the other models for short-term maturities. Reasons for this might be that LSTM models excel at learning from historical sequences and detecting hidden patterns that may not be apparent in linear models like VAR or the RW model. The memory and non-linear activation functions of an LSTM can potentially capture long-term dependencies and adapt to changing market conditions better, whereas traditional models may fall short in capturing the underlying dynamics effectively.

Table 7.3: Implied Volatility Out-of-Sample Average Prediction Errors by Moneyness and Maturity

Comparison of out-of-sample prediction errors using a variety of models, expressed as average of the three performance measures, i.e. Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean Directional Error (MDE). We consider the implied volatility surface backed out of the predicted beta coefficients for the period from August 2022 to December 2022 on an eight hour frequency and divide the results for different maturities (Short Term: 3-14 days; Mid Term: 14-30 days; Long-Term: 30-270 days), moneyness levels (OTM Put: moneyness between 0.5-0.94; ATM: moneyness between 0.94-1.06; OTM Call: moneyness between 1.06-2) as well as any permutation of these two. The last row and column depict the overall prediction errors. We highlight the overall best results in blue.

	Short-Term			Mid-Term			Long-Term			Total		
	RMSE	**MAE**	**MCD**	**RMSE**	**MAE**	**MCD**	**RMSE**	**MAE**	**MCD**	**RMSE**	**MAE**	**MCD**
OTM Put:												
RW	0.08	0.0485	50.3%	0.066	0.0442	50.72%	0.0884	0.0513	51.19%	0.0838	0.05	51.49%
VAR	0.0786	0.0453	53.37%	0.0624	0.0417	52.78%	0.0389	0.028	54.01%	0.05	0.0326	53.7%
LSTM	0.0664	0.0418	62.39%	0.0604	0.0411	61.48%	0.0459	0.0352	53.92%	0.0516	0.0371	56.12%
ATM:												
RW	0.0454	0.0339	51.64%	0.0375	0.0266	45.01%	0.0768	0,0389	50.57%	0.0614	0.0346	50.59%
VAR	0.0429	0.033	53.86%	0.0327	0.0241	48.72%	0.0246	0.0176	51.43%	0.0326	0.0235	51.65%
LSTM	0.0383	0.0294	64.63%	0.0304	0.0229	59.59%	0.033	0.025	52.62%	0.034	0.0258	57.66%
OTM Call:												
RW	0.0536	0.0377	48.51%	0.0447	0.0308	45.96%	0.0811	0.0431	51.39%	0.0763	0.0414	51.06%
VAR	0.0484	0.0359	49.83%	0.0403	0.029	46.31%	0.0289	0.0212	50.72%	0.0321	0.0231	50.18%
LSTM	0.0397	0.031	62.1%	0.0365	0.0266	59.02%	0.036	0.0277	52.84%	0.0363	0.0279	53.99%
Total:												
RW	0.0621	0.0403	50.27%	0.0532	0.0353	47.7%	0.0832	0.045	51.48%	0.0868	0.0474	50.79%
VAR	0.0595	0.0382	52.55%	0.0493	0.0331	49.55%	0.0323	0.0231	51.91%	0.0397	0.0268	51.83%
LSTM	0.0505	0.0343	63.11%	0.0468	0.0317	60.17%	0.0394	0.03	53.19%	0.0421	0.0305	55.32%

Table 7.4: Diebold-Mariano Test Statistic

Pairwise comparison of the Diebold and Mariano (1995) (DM) statistic for out-of-sample predictions from August 2022 to December 2022. We divide the results for different maturities (Short Term: 3-14 days; Mid Term: 14-30 days; Long-Term: 30-270 days), moneyness levels (OTM Put: moneyness between 0.5-0.94; ATM: moneyness between 0.94-1.06; OTM Call: moneyness between 1.06-2) as well as any permutation of these two. The DM test compares the forecasting accuracy of the RW/VAR (column) with the VAR/LSTM (row) model. It tests the null hypothesis that the forecasts have the same accuracy, i.e. a positive DM statistic indicates that the first model has larger average forecast error compared to the second and hence the second model is preferred. The p-value of the statistic is expressed using one (three) asterisks (***) for the 10% (1%) significance level.

	Short-Term		Mid-Term		Long-Term		Total	
	VAR	**LSTM**	**VAR**	**LSTM**	**VAR**	**LSTM**	**VAR**	**LSTM**
OTM Put:								
RW	0.44	3.48^{***}	2.77^{***}	2.83^{***}	21.19^{***}	15.86^{***}	11.03^{***}	10.48^{***}
VAR		4.31^{***}		1.07		-13.93^{***}		1.38
ATM:								
RW	5.04^{***}	10.69^{***}	8.15^{***}	7.4^{***}	12.71^{***}	9.66^{***}	15.33^{***}	15.44^{***}
VAR		10.41^{***}		3.64^{***}		-9.1^{***}		4.26^{***}
OTM Call:								
RW	6.32^{***}	8.04^{***}	7.01^{***}	5.31^{***}	25.95^{***}	20.37^{***}	27.27^{***}	22.35^{***}
VAR		6.94^{***}		3.07		-15.5^{***}		-4.46^{***}
Total:								
RW	1.84^{*}	6.64^{***}	6.06^{***}	6.21^{***}	36.55^{***}	33.31^{***}	36.6^{***}	34.4^{***}
VAR		7.07^{***}		2.67^{***}		-43.97^{***}		-9.92^{***}

7.3.2 Economic Value

The findings in Table 7.3 support the choice to select only short-term ATM options for the trading evaluation. Within this subgroup, we face up to two actively traded maturities, each consisting of two to four strike levels. At each time we invest \$1000 in the hypothetical portfolios A, B and C, and document the return, as described in Subsection 7.2.3. Note that strategy A considers only the directional forecast of each model whereas B accounts for each exact prediction throughout the options chain. Table 7.5 summaries the trading performance of the strategies using the three model predictions over the out-of-sample period from August to December 2022. It presents the percentage mean, median and standard deviation of the eight-hour returns as well as the Sharpe- and Sortino-ratio.

The first strategy reveals that only the LSTM forecasts procdue a positive mean return of 1.18%. In contrast, both the RW and VAR models incur losses on average, with returns of -0.18% and -0.12%, respectively. This can be attributed to their weak predictive power for direction. Recall that both RW and VAR model can predict the correct direction in merely

50% of the time, while the LSTM is boasts a 64% success rate. Accurate directional prediction is of utmost importance for a straddle strategy, especially considering bitcoin's proneness for sudden price jumps. The marginally better results for the VAR model, compared to the RW model, further confirm the findings in Table 7.3. Moreover, the LSTM model provides the lowest standard deviation, resulting in impressive Sharpe and Sortino ratios. However, each trade requires the purchase (sale) of an options portfolio and exit the position, forcing the trader to cross the spread. With an average spread of 3.2% for ATM bitcoin options, this presumably profitable straddle strategy would result in losses.

We find similar results for the second strategy. Once again, the only profitable trading strategy is based on the LSTM forecasts, with an average return of 2.34%. The other two models lag behind considerably, with -1.45% and -1.74% returns, respectively. However, we face an entirely different situation now. Note that we use the predicted IV to calculate each specific delta along the options chain and use this information to delta-hedge or option position. A closer examination of the data reveals that the forecasted IVs occasionally deviate significantly from the true value, particularly for options farther from ATM. Given the short time to maturity, the second closest option to ATM might still exhibit a 30 delta which is considered OTM for other asset classes. In fact, it is a combination of incorrect predictions and a low initial portfolio value V_t^B that produces extreme positive or negative returns on investment, explaining the high standard deviation but low median of the second trading strategy. That is, we have observed numerous occasions where the portfolio value V_t^B is almost offset, i.e. the delta-hedged positions in the portfolio cancel each other out. However, the wrong delta calculated using the predictions results in an offsetting options position but great profit/loss on the futures leg, implying profits (losses) of up to 250% (-190%). Both the Sharpe and Sortino ratios reflect this IV misprediction. Although the average return for the delta-hedged portfolio surpasses that of the straddle strategy by more than double, the Sharpe and Sortino ratios are are more in favor for the straddle, suggesting a preference for strategy A over B.

Interestingly, the 'hodl' benchmark (C) displays the highest standard deviation across all models and strategies. In fact, it ranks as the third worst model after the delta-hedged RW and VAR in terms of mean return, Sharpe, and Sortino ratios. However, the spread for the underlying spot or futures is generally negligible, rendering the buy-and-hold strategy not substantially worse than the others.

Table 7.5: Trading Strategy Evaluation

Summary statistics of the three trading strategies (straddle, delta-hedge and buy-and-hold) for short-term ATM options based on trading signals generated by distinct predictive models (RW, VAR and LSTM). The performance measures include the average return, standard deviation, resulting Sharpe Ratio, and Sortino Ratio, all expressed in percentage terms for the out-of-sample period August to December 2022 using the intraday eight hour rebalancing frequency. We have highlighted the best model in blue.

	Mean	Std. Dev.	Median	Sharpe Ratio	Sortino Ratio
Strategy A *(ATM Straddle)*					
RW	-0.18	6.31	-0.76	-2.94	-4.46
VAR	-0.12	6.14	-0.96	-2.03	-3.3
LSTM	1.18	5.85	0.52	20.13	38.56
Strategy B *(delta-hedged portfolio)*					
RW	-1.45	19.15	-0.15	-7.55	-9.94
VAR	-1.74	15.56	-0.21	-11.14	-13.34
LSTM	2.34	19.33	0.22	12.09	20.78
Strategy C *(Buy and Hodl)*	-0.94	28.1	-1.07	-4.54	-5.6

7.4 Concluding Remarks

This chapter introduces and implements a technical framework capable of capturing and modeling the dynamics of bitcoin implied volatility in a manner that enables traders and portfolio managers to generate positive returns. Our approach is two-fold: first, we model the implied surface with a parametric model; and second, we use the resulting parameters to create a time series, modeling and predicting these using both statistical and machine learning models before benchmarking the predictions against a naive forecast. For the former, we advocate the use of a second-order Taylor expansion to capture the key components driving the surface (Dumas et al., 1998). This approach is easy to implement, intuitive for traders, and capable of capturing the surface to a satisfactory extent, as we demonstrate using an array of statistical evaluation measures like AIC, BIC, (adjusted) R^2 or RMSE.

With regard to time series modeling, we compared three distinct approaches to obtain step-ahead predictions: a random walk, VAR, and LSTM model, each with optimized lag. Given the predicted model parameters, we translate these back into implied volatilities and compare their accuracy with the actual observed market implied volatilities, measuring their performance based on common statistical elements like RMSE, MAE, and MDE. Motivated by the discussion in Section 3.2, we partition the surface into various maturity and moneyness subcategories and find clear differences between groups. While the best overall results are obtained from a VAR model concerning the RMSE and MAE, we find rather poor directional

forecasts. Nevertheless, using the LSTM on short-term ATM options, we are able to predict the right direction of the implied volatility in over 60% of the cases. Considering that this subgroup contains the most overall trading volume, we decide to focus further research on short-term ATM options only, i.e., those expiring within the next two weeks.

Based on this setup, we have introduced three distinct trading strategies and evaluated their profitability. Namely, we consider a ATM straddle, delta-hedged portfolio and a simply buy-and-hold strategy. For any trading signal obtained from the model, we either buy or sell the respective option and re-evaluate our position at the next time increment. We backtest these strategies on an out-of-sample period from August to December 2022 and find mixed results. The VAR and RW model are not able to generate forecasts sufficient enough to generate positive returns for any of the tested strategies. On the other hand, the LSTM forecasts yield – on average – positive returns for both the delta-hedge and ATM straddle strategy. Although the delta-hedged approach yields a higher average return, its standard deviation greatly exceeds that of the ATM straddle, and consequently the delta-hedged Sharpe and Sortino ratio are worse.[17] Contrary, the ATM straddle shows fantastic results for both Sharp and Sortino ratio.

While only the LSTM strategy-model pairs yielded positive average returns, any strategy failed to be implemented into real market conditions due to transaction costs and fees. This is a strong indicator of an efficient Bitcoin options market. Furthermore, we have shown that Monte Carlo Dropout opens the door to an entirely new option pricing research strain. Despite being an effective model that showed promising results, our LSTM model would benefit from a larger historical dataset. Similar to the volatility index, this work is ahead of its time as we are only scratching the surface of the possibilities of ML in financial data. However, we believe that in 2-3 years' time, there will be sufficient market data to train regimes and dynamics more accurately. Moreover, the model could include more external information in the form of another time series, such as the correlation between the implied volatility and the spot, or the open interest of call and put options. A hybrid model between LSTM and VAR could also be something to consider. Finally, it would be interesting to investigate a threshold for the size of the signal, i.e. buy/sell an option only if a certain signal level is exceeded.

Overall, this chapter has contributed to understanding bitcoin implied volatility dynamics and its efficiency, and provided insights into potential profitable trading strategies. Future research could extend this work by incorporating additional information and more advanced modeling techniques to further improve the performance of the trading strategies.

---

[17]Note that the ATM straddle requires a directional forecast only, while the delta-hedged approach considers the size of the implied volatility to calculate the delta. In fact, in some occasions the forecasts were far too off from the true value which resulted in abnormal returns for the delta-hedge strategy and consequently the higher standard deviation.

CHAPTER 8

CONCLUSION

The internet as we know it today is – in its core – nothing more than a string of different protocols, including but not limited to IP or HTTP etc. Blockchains are basically no different and operate in similar spirit. They use mostly fully automated protocols, i.e. smart contracts, which run in the background and serve some purpose like providing access to or storing data, settle transactions or transfer value. A distinguishing feature of these technologies, however, lies in their decentralised architecture and the resulting transparency. Hence, the question is not 'if', but 'when' these two different types of protocols will interlink and blockchain projects will (unconsciously) find their way into the everyday life of the general population.[1]

In this dissertation, we operate on the forefront of this transition and provide fundamental research for what is to become "the next big thing" in the field of cryptocurrencies. We focus on trading bitcoin options and its implied volatility extensively, thus establishing the basis for further research in this area. Since its introduction in January 2009, bitcoin consistently had the most trading volume, most media coverage and most controversy surrounding it among all digital assets. No cryptocurrency has been declared dead more often, only to reach new all-time heights. This has resulted in great interest for bitcoin derivatives, particularly within the options market. This market has been steadily growing since its inception in 2017, enabling agents to trade the very essence of bitcoin, i.e. its volatility. Our research shows a considerable increase in the trading volume of bitcoin options, far outstripping the growth observed for S&P 500 options, alongside the institutional interest in the bitcoin option market. All these factors contributed to an increase in liquidity and declining bid-ask spread, thereby intensifying the competition to uphold an edge. Furthermore, we identify clusters in trading activity and assign these to potential groups of retail or institutional traders.

---

[1]One currently often mentioned and discussed topic are Central Bank Digital Currencies (CBDC). The idea is rather simple: Any central banks runs its private blockchain and issues digital token worth one USD/GBP/EUR. The central banks claim to have a better overview about current circulation and spending while critics see this as a step closer to total surveillance.

However, the fragmentation of the crypto derivatives market and idle behaviour of regulators created an array of different option settlement mechanisms. In the first research chapter, we present and examine all available option types currently tradable on regulated and unregulated exchanges. We derive the mathematical pricing formulae under the assumption of a Black and Scholes (1973) framework, and first show that inverse options, which account for approximately 90% of the overall traded volume and open interest, can and should be priced according to Garman and Kohlhagen (1983). Furthermore, we point out the market incompleteness for the biggest crypto options exchanges and highlight key risks associated with stablecoin-denominated options which represent all other tradable options on unregulated exchanges. We emphasises the use of quanto options to tackle the de-pegging risk of stablecoins and introduce an attractive, yet affordable exotic alternative to inverse options, i.e quanto inverse, to enable wider participation in the bitcoin options market. Both quanto and quanto inverse options are natural extensions of the currently available products and ensure investor protection and increase participation in the market.

Progressing with our empirical research, we conclude that bitcoin implied volatility makes a poor hedge for the underlying. Our analysis shows a frequent fluctuation in the correlation between the bitcoin's price and its implied volatility, in both magnitude and sign. Consequently, the leverage effect, which we find in most traditional markets, does not hold for bitcoin, indicating similarities to FX implied volatilities. This requires traders to reassess their hedged options position and reevaluate their delta. In fact, we find evidence for a clear regime-dependency of the spot-volatility correlation. However, bitcoin implied volatility also exhibits characteristics akin to other asset classes like equities or commodities. For instance, sudden jumps in the underlying make it spike up before it corrects itself in a mean-reverting manner with both backwardation and contango periods clearly visible. As one might anticipate, the overall level of implied volatility is significantly higher than in other asset classes, although we do find instances where commodities depict higher volatility. A distinctive characteristic of bitcoin implied volatility is its tail behaviour. While the standard implied volatility resembles a hockey-stick shaped smirk, i.e. OTM put exceeding OTM calls with a local minima around ATM levels, we encounter in many instances a strong variation in shape, i.e. symmetric or even positive skew in which the OTM call exceeds the OTM put implied volatility.

These findings motivate our second research chapter which examines the dynamic hedging potential of various model-free smile-implied and regime-dependent smile-adjusted delta for bitcoin options with different maturities and moneyness levels. For this, we consider the unique derivatives of the crypto market and compare two distinct hedging instruments, i.e. perpetuals and fixed-maturity futures, at an intra-day rebalancing frequency. Previous empirical academic research on this strain of has only investigated equity index options. Our study prioritise the robust, model-free framework that is preferred by so many practitioners. While our findings are somewhat mixed, the general consensus of conclusions is that smile-adjusted hedge ratios can only improve on the BS delta for OTM put options, sometimes. In some

instances, we have demonstrated that the smile-implied (sticky-moneyness) delta can offer a substantially superior hedge than a standard BS delta for OTM options, with efficiency gains of over 40% in some cases. Furthermore, we find that bid-ask spreads on bitcoin options are still relatively large, particularly at the tails, and much higher than they are on bitcoin futures or perpetuals. Hence, the profitability of bitcoin options market making hinges more on accurate dynamic delta-hedging than delta-gamma-vega hedging. Nevertheless, if spreads on bitcoin options were to reduce in future, it would be interesting to investigate gamma and vega hedging for a bitcoin options book or even different underlyings like ether or solana. However, as of this writing, the trading costs of hedging price and volatility risk with options could potentially offset any additional profits derived from possibly increasing trading volumes due to a reduction in spreads

Beyond simply taking a directional (hedged) position in the underlying, our study illustrates how bitcoin options can be used to generate implied volatility indices across various maturities. These indices consider the unique trading patterns of bitcoin options and serve as a strike price for variance swaps. In a pioneering effort, we analyse the time series of the bitcoin implied volatility index and emphasise its use as settlement value for futures trading. Our findings confirm consistency of the bitcoin implied volatility index with those of other classes, i.e. we identify extensive periods in contago and abrupt cycles in backwardation after price crashes. Further research reveals that the bitcoin carry, i.e. the difference between implied and realised volatility, is typically very high, reaching levels up to 30% which indicates significant overpricing of bitcoin options. Following this line of inquiry, we regard the calculated value of the bitcoin implied volatility index as a fair swap rate for a variance swap, allowing us to examine its variance risk premium, i.e. the payoff to a variance swap. In an attempt to position bitcoin on the global market better, we compare its VRP to other asset classes like US equity, commodities and FX rates. Interestingly, our findings demonstrate that bitcoin's VRP mimics those of other asset classes, typically exhibiting a negative trend with sudden surges into positive territory, albeit with much larger magnitudes. Despite its tainted image, our comparison reveals that bitcoin has the most appealing VRP profile for market makers, making it an attractive asset to trade. An examination of the determinants of bitcoin volatility and its correlation to other assets yields no significant results. Indeed, in recent years, bitcoin appears to have lost its diversification potential, a development we mainly attribute to its increasing correlation with other major markets. Future research could potentially include the index and resulting VRP behaviour of other coins and token like ether or solana. Given the increase of trading volume for ether options in the last year, it is plausible to assume that one could create an ETH implied volatility index, while it may still be premature for solana. However, with the launch of futures on the DVOL, new opportunities have emerged for researching the pricing and modelling of bitcoin VIX derivatives.

Our research has thus far focused primarily on slices of the implied surface, i.e. skews with fixed maturity. In other words, we neglect most of the entire surface. In our final

research chapter we aim to model and predict the dynamics of the bitcoin implied volatility surface. For this, we pursue a two-step approach examining two main questions. First, how well does out model replicate the surface; and second, how well can the time series models predict the next step. In addressing the former, we consider a model-free approach in which we fit intra-day market implied volatility surface data to an array of parametric second-order Taylor expansions (Dumas et al., 1998). We identify a five-parameter model that most closely replicates the surfaces and ensure its arbitrage-free nature. However, due to option clusters for short-term ATM, it is challenging to capture tails with the second-order polynomial. Hence, we find the most accurate results for short-term ATM IV. Next, we model the rustling parameter time series. Motivated by the serial autocorrelated properties of the time series, we fit a VAR model to predict the next steps. In light of the growing interest in ML models, we integrate a LSTM model to our portfolio of predictive models, as a natural non-linear extension of the VAR. Given the statistical success in predicting the volatilities, we examine the economic value of our results through different trading strategies. Interesting, this rudimentary approach yields encouraging results for short-term ATM options as the trading strategies yield a positive mean return and high Sharpe and Sortino ratios when backtesting on out-of-sample data. Our findings confirm the results of Bolch and Book (2022) on S&P 500 options, i.e. short-term ATM options are predictable and it is possible to exploit these. However, the practical implications of real-world transaction costs or bid-ask spreads effectively nullify these results, indicating an efficient bitcoin options market.

The potential for further research in this domain are immense. Firstly, it would be interesting to consider a dynamic regime-switching model to capture volatility regimes, e.g. Hamilton's Markos Switching model. In the same vein, a dynamic mixture model combining VAR and LSTM might outperform a single model. Additionally, incorporating the spot-vol correlation or changes in the open interest for the moneyness-maturity subcategories may enhance forecast accuracy. On the other side, we are only scratching on the surface of ML application on financial data. The implemented Monte Carlo Dropout function paves the way to an entirely new research strain of pricing and hedging options as the MCDO provides a probability distribution of future option or underlying prices. In general, our approach could be readily extended to any other statistical or ML model, such as HAR, convolutional LSTM, or Transformer models, which might yield improved accuracy. However, the currently available data is insufficient to apply these models just yet. Similar to previous research, the predictability of the ETH implied volatility surface, or joint dynamic modelling, would significantly enhance our understanding of the cryptocurrency options market. Finally, it would be interesting to investigate whether a threshold for the signal would yield improved results.

In many aspects, this thesis is ahead of its time. Often mis-perceived as a niche market, we are the first to provide fundamental research for these promising crypto derivatives. The presented methodologies, if not already, can and should be used to model, price and hedge bitcoin options and their implied volatility. Our proprietary dataset is unmatched in cur-

rent literature and covers the longest period of historical bitcoin option market data. Thus, we examine existing products and those not yet available, or only traded OTC, and provide valuable insight into the bitcoin options market for both academics and practitioners. The continuous inflow of institutional players has transformed the bitcoin options market considerably over the past three years. As a result, we have seen a rise in correlation between bitcoin and large equity indices like S&P 500. Further influx of institutional liquidity will most likely moderate the volatile nature of cryptocurrencies. However, prior to that, international regulation will have to take place to provide a foundation on which trust in this asset class can be built. We believe that bitcoin, as the prime cryptocurrency, will prevail as a highly traded and liquid asset, but it will not lead the inevitable changes of the financial landscape. Its architecture is too ancient and ill-equipped for modern tasks. Nevertheless, its influence over other cryptocurrencies is indisputably strong, and as such, bitcoin can be seen as a proxy for the overall cryptocurrency state. The findings of this dissertation help both scholars and practitioners understand the cryptocurrency derivatives market better and pave the way for further research as they allow other academics to build on our analyses.

BIBLIOGRAPHY

Aalborg, H., P. Molnár, and F. de Vries (2019). What can explain the price, volatility and trading volume of bitcoin? *Finance Research Letters 29*, 255–265.

Akyildirim, E., S. Corbet, B. Lucey, A. Sensoy, and L. Yarovaya (2020). The relationship between implied volatility and cryptocurrency returns. *Finance Research Letters 33*, 101212.

Alexander, C. (2000). Principal component analysis of volatility smiles and skews. *SSRN Working Paper*.

Alexander, C. (2008). *Pricing, Hedging and Trading Financial Instruments*. Market Risk Analysis III. Wiley.

Alexander, C., D. Chen, and A. Imeraj (2023). Crypto quanto and inverse options. *Mathematical Finance 33*(4), 1005–1043.

Alexander, C., J. Choi, H. Park, and S. Sohn (2020). Bitmex bitcoin derivatives: Price discovery, informational efficiency, and hedging effectiveness. *Journal of Futures Markets 40*(1), 23–43.

Alexander, C., J. Deng, J. Feng, and H. Wan (2022). Net buying pressure and the information in bitcoin option trades. *Journal of Financial Markets* (100764).

Alexander, C., J. Deng, and B. Zou (2023). Hedging with automatic liquidation and leverage selection on bitcoin futures. *European Journal of Operational Research 306*(1), 478–493.

Alexander, C. and A. Imeraj (2021). The bitcoin vix and its variance risk premium. *The Journal of Alternative Investments 23*(4), 84–109.

Alexander, C. and A. Imeraj (2023). Delta hedging bitcoin options with a smile. *Quantitative Finance 23*(5), 799–817.

Alexander, C. and D. Korovilas (2013). Volatility exchange-traded notes: Curse or cure? *Journal of Alternative Investments 15*(2), 52–70.

Alexander, C. and L. Nogueira (2007a). Model-free hedge ratios and scale-invariant models. *Journal of Banking and Finance 31*(6), 1839–1861.

Alexander, C. and L. Nogueira (2007b). Model-free price hedge ratios for homogeneous claims on tradable assets. *Quantitative Finance 7*(5), 473–479.

Alexander, C. and J. Rauch (2021). A general property for time aggregation. *European Journal of Operational Research 291*(2), 536–548.

Alexander, C., A. Rubinov, M. Kalepky, and S. Leontsinis (2012). Regime-dependent smile-adjusted delta hedging. *Journal of Futures Markets 32*(3), 203–229.

Ammann, M. and R. Buesser (2013). Variance risk premiums in foreign exchange markets. *Journal of Empirical Finance 23*, 16–32.

Ankirchner, S., P. Imkeller, and G. Dos Reis (2010). Pricing and hedging of derivatives based on nontradable underlyings. *Mathematical Finance 20*(2), 289–312.

Attie, L. (2017). The performance of smile-implied delta hedging. *Canadian Derivatives Institute, Technical Note TN 17-01*.

Audrino, F. and D. Colangelo (2009). Semi-parametric forecasts of the implied volatility surface using regression trees. *Statistics and Computing 20*, 421–434.

Aït-Sahalia, A., M. Karaman, and L. Mancini (2020). The term structure of equity and variance risk premia. *Journal of Econometrics 219*(2), 204–230.

Aït-Sahalia, Y. (2004). Disentangling diffusion from jumps. *Journal of Financial Economics 74*(3), 487–528.

Aït-Sahalia, Y. and A. Lo (2002). Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance 53*(2), 499–547.

Badshah, I., B. Frijns, and A. Tourani-Rad (2013). Contemporaneous spill-over among equity, gold, and exchange rate implied volatility indices. *Journal of Futures Markets 33*(6), 555–572.

Bakshi, G., C. Cao, and Z. Chen (1997). Empirical performance of alternative option pricingmodels. *Journal of Finance 52*, 2003–2049.

Bakshi, G. and N. Kapadia (2003). Delta-hedged gains and the negative market volatility risk premium. *Review of Financial Studies 16*(2), 527–566.

Bakshi, G., N. Kapadia, and D. Madan (2003). Stock return characteristics,skew laws, and the differential pricing of individual equity options. *Review of Financial Studies 16*(1), 101–143.

Bakshi, G. and D. Madan (2006). A theory of volatility spreads. *Management Science 52*(12), 1945–1956.

Bandi, F., N. Fusari, and R. Renô (2021). Structural stochastic volatility. *SSRN Working Paper*.

Bandi, F. and R. Reno (2016). Price and volatility co-jumps. *Journal of Financial Economics 119*(1), 107–146.

Barndorff-Nielsen, O. and N. Shephard (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics 2*(1), 1–37.

Barndorff-Nielsen, O. and N. Shephard (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics 4*(1), 1–30.

Bates, D. (1996). Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *Review of Financial Studies 9*(1), 169–107.

Bates, D. (2005). Hedging the smirk. *Finance Research Letters 2*(4), 195–200.

Baur, D. and L. Hoang (2021). A crypto safe haven against bitcoin. *Finance Research Letters 38*.

Baur, D. G. and T. Dimpfl (2018). Asymmetric volatility in cryptocurrencies. *Economics Letters 173*, 148–151.

Bedendo, M. and S. Hodges (2009). The dynamics of the volatility skew: A kalman filter approach. *Journal of Banking and Finance 33*, 1156–1165.

Beer, S. and H. Fink (2019). Dynamics of foreign exchange implied volatility and implied correlation surfaces. *Quantitative Finance 19*(8), 1293–1320.

Bekaert, G. and M. Hoerova (2014). The vix, the variance premium and stock market volatility. *Journal of Econometrics 183*, 181–192.

Bekaert, G., M. Hoerova, and M. Ducab (2013). Risk, uncertainty and monetary policy. *Journal of Monetary Economics 60*(7), 181–192.

Ben Cheikh, N., Y. Ben Zaied, and J. Chevallier (2020). Asymmetric volatility in cryptocurrency markets: New evidence from smooth transition garch models. *Finance Research Letters 35*, 101293.

Bennett, C. (2014). *Trading volatility, correlation, term structure and skew*. Independent.

Bernales, A. and M. Guidolin (2014). Can we forecast the implied volatility surface dynamics of equity options? predictability and economic value tests. *Journal of Banking and Finance 46*(1), 326–342.

Bernard, C. and Z. Cui (2014). Prices and asymptotics for discrete variance swaps. *Journal of Applied Mathematical Finance 21*(2), 140–173.

Bernard, C., Z. Cui, and D. McLeish (2014). Convergence of the discrete variance swap in time-homogeneous diffusion models. *Quantitative Finance Letters 2*(1), 1–6.

Björk, T. (2020). *Arbitrage Theory in Continous Times*. Okford Finance.

Björk, T. and I. Slinko (2006). Towards a general theory of good deal bounds. *Review of Finance 10*(2), 221–260.

Black, F. (1976). The pricing of commodity contracts. *Journal of Financial Economics 3*, 231–237.

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

Bliss, R. and N. Panigirtzoglou (2002). Testing the stability of implied probability density functions. *Journal of Banking and Finance 26*(2-3), 381–422.

Bolch, D. and A. Book (2022). Smiling in action. *Wilmott July*(120), 70–81.

Bollerslev, T., G. Tauchen, and H. Zhou (2009). Expected stock returns and variance risk premia. *Review of Finance 22*(11), 4463–4492.

Bollerslev, T. and V. Todorov (2011). Tails, fears, and risk premia. *Journal of Finance 66*(6), 2165–2211.

Bondarenko, O. (2003). Estimation of risk-neutral densities using positive convolution approximation. *Journal of Econometrics 116*(1), 85–112.

Bondarenko, O. (2014). Variance trading and market price of variance risk. *Journal of Econometrics 180*(1), 81–97.

Bouchaud, J.-P., A. Matacz, and M. Potters (2001). Leverage effect in financial markets: The retarded volatility model. *Physical Review Letters 87*(22), 228701/1–228701/4.

Bouri, E., D. Roubaud, and S. Shahzad (2020). Do bitcoin and other cryptocurrencies jump together? *Quarterly Review of Economics and Finance 76*, 396 – 409.

Box, G. and M. Gwilym (1970). *Time Series Analysis: Forecasting and Control*. Wiley.

Brandt, M. W. and T. Wu (2002). Cross-sectional tests of deterministic volatility functions. *Journal of Empirical Finance 9*(5), 525–550.

Breeden, D. and R. Litzenberger (1978). Prices of state contingent claims implicit in option prices. *Journal of Business 51*, 621–651.

Brenner, M. and D. Galai (1989). New financial instruments for hedge changes in volatility. *Financial Analysts Journal 45*(4), 61–65.

Brigo, D. and F. Mercurio (2002). Lognormal-mixture dynamics and calibration to market volatility smiles. *International Journal of Theoretical and Applied Finance 5*(4), 427–446.

Britten-Jones, M. and A. Neuberger (2000). Option prices, implied price processes, and stochastic volatility. *Journal of Finance 55*(2), 839–866.

Broadie, M., M. Chernov, and M. Johannes (2007). Model specification and risk premia: Evidence from futures options. *Journal of Finance 62*(3), 1453–1490.

Broadie, M. and A. Jain (2008a). The effect of jumps and discrete sampling on volatility and variance swaps. *International Journal of Theoretical and Applied Finance 11*(8), 761–797.

Broadie, M. and A. Jain (2008b). Pricing and hedging volatility derivatives. *Journal of Derivatives 15*(3), 7–24.

Broomhead, D. and D. Lowe (1988). Multi-variable functional interpolation and adaptive networks. *Complex Systems 2*, 327–355.

Brownlee, J. (2021). *Deep Learning for Time Series Forecasting*. Jason Brownlee.

Buehler, H., L. Gonon, J. Teichmann, and B. Wood (2019). Deep hedging. *Quantitative Finance 19*(8), 1271–1291.

Cao, J., J. Chen, and J. Hull (2020). A neural network approach to understanding implied volatility movements. *Quantitative Finance 20*(9), 1405–1413.

Cao, M. and B. Celik (2021). Valuation of bitcoin options. *Journal of Futures Markets 41*(7), 1007–1026.

Carr, P., H. Geman, D. Madan, and M. Yor (2002). The fine structure of asset returns: An empirical investigation. *The Journal of Business 75*(2), 305–332.

Carr, P. and R. Lee (2009a). Robust replication of volatility derivatives. *University of Chicago Working Paper*.

Carr, P. and R. Lee (2009b). Volatility derivatives. *Annual Review of Financial Economics 1*(1), 319–339.

Carr, P. and D. Madan (1999). Option valuation using the fast fourier transform. *Journal of Computational Finance 2*, 61–73.

Carr, P. and D. Madan (2001). *Towards a theory of volatility trading*. Handbooks in Mathematical Finance. Cambridge University Press.

Carr, P. and D. Madan (2005). A note on sufficient conditions for noarbitrage. *Finance Research Letters 2*(3), 125–130.

Carr, P. and L. Wu (2004). A tale of two indices. *Baruch College Working Paper*.

Carr, P. and L. Wu (2007). Stochastic skew in currency options. *Journal of Financial Economics 86*(1), 213–247.

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

Cauchy, A. (1847). Méthode générale pour la résolution de systèmes d'équations simultanées. *Compte rendu des séances de l'académie des sciences*, 536–538.

CBOE Global Markets, I. (2003). CBOE VIX white paper. *Available from CBOE website*.

Chaim, P. and M. Laurini (2018). Volatility and return jumps in bitcoin. *Economics Letters 173*, 158–163.

Chalamandaris, G. and A. Tsekrekos (2010). Predictable dynamics in implied volatility surfaces from otc currency options. *Journal of Banking and Finance 34*(6), 1175–1188.

Chen, J., M. Clements, and A. Urquhart (2022). Forecasting bitcoin. *SSRN Working Paper*.

Chen, K.-S. and Y.-C. Huang (2021). Detecting jump risk and jump-diffusion model for bitcoin options pricing and hedging. *Mathematics 9*(20).

Chen, S. and Z. Zhang (2019). Forecasting implied volatility smile surface via deep learning and attention mechanism. *arXiv*.

Cheung, A., E. Roca, and J. SU (2015). Crypto-currency bubbles: An application of the Phillips – Shi – Yu methodology on Mt. Gox bitcoin prices. *Applied Economics 47*(23).

Chi, Y. and W. Hao (2021). A Horserace of Volatility Models for Cryptocurrency: Evidence from Bitcoin Spot and Option Markets. *Journal of International Financial Markets, Institutions and Money 75*.

Cho, K., B. van Merriënboer, D. Bahdanau, and Y. Bengio (2014). On the properties of neural machine translation: Encoder–decoder approaches. *Proceedings of SSST 2014 - 8th Workshop on Syntax, Semantics and Structure in Statistical Translation*, 103–111.

Choi, H., P. Mueller, and A. Vedolin (2017). Bond variance risk premiums. *Review of Finance 2*(3), 987–1022.

Chow, K. V., W. Jiang, and J. Li (2021). *Does VIX Truly Measure Return Volatility?* Handbook of Financial Econometrics, Mathematics, Statistics, and Machine Learning. World Scientific.

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

Clark, I. (2011). *Foreign Exchange Option Pricing: A Practitioner's Guide*. Wiley Finance Series.

Coleman, T., Y. Kim, Y. Li, and A. Verma (2001). Dynamic hedging with deterministic local volatility function model. *Journal of Risk 4*(1), 63–89.

Cont, R. and J. Da Fonseca (2002). Dynamics of implied volatility surfaces. *Quantitative Finance 2*(1), 45–60.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics 7*(2), 174–196.

Cox, J. (1975). Notes on option pricing i: Constant elasticity of diffusions. *Stanford University Working Paper*.

Cox, J., S. Ross, and M. Rubinstein (1979). Option pricing: A simplified approach. *Journal of Financial Economics 7*(3), 229–263.

Cretarola, A., G. Figà-Talamanca, and M. Patacca (2020). Market attention and bitcoin price modeling: theory, estimation and option pricing. *Decisions in Economics and Finance 43*(1), 187–228.

Crépey, S. (2004). Delta-Hedging Vega Risk? *Quantitative Finance 4*(5), 559–579.

Daglish, T., J. Hull, and W. Suo (2007). Volatility surfaces: theory, rules of thumb, and empirical evidence. *Quantitative Finance 7*(5), 507–524.

D'Amato, V., S. Levantesi, and G. Piscopo (2022). Deep learning in predicting cryptocurrency volatility. *Physica A: Statistical Mechanics and its Applications 596*.

Davis, M. (2000). Optimal hedging with basis risk. *Financial and Actuarial Mathematics Group Working Paper, TU Vienna*.

De Spiegeleer, J., D. Madan, S. Reyners, and W. Schoutens (2018). Machine learning for quantitative finance: fast derivative pricing, hedging and fitting. *Quantiative Finance 18*(10), 1635–1643.

Demeterfi, K. (1998). How to value and hedge options on foreign indexes. *Quantitative Strategies Research Notes*, 1–26.

Demeterfi, K., E. Derman, M. Kamal, and J. Zou (1999). A guide to volatility and variance swaps. *The Journal of Derivatives 6*(4), 9–32.

Deng, J., H. Pan, S. Zhang, and B. Zou (2020). Minimum-variance hedging of bitcoin inverse futures. *Applied Economics*, 6320–6337.

Deng, J., H. Pan, S. Zhang, and B. Zou (2021). Optimal bitcoin trading with inverse futures. *Annals of Operations Research 304*(1-2), 139–163.

Derman, E. (1999). Regimes of volatility. *Risk 12*(4), 55–59.

Derman, E., M. Kamal, I. Kani, J. McClure, C. Pirasteh, and J. Zou (1996). Investing in volatility. *Quantiative Strategies Research Notes*.

Derman, E. and Kani (1994). The volatility smile and its implied tree. *Quantitative Strategies Research Notes*.

Derman, E., I. Kani, and J. Zou (1996). The local volatility surface: Unlocking the information in index option prices. *Financial Analysts Journal 52*, 25–36.

Diebold, F. and R. Mariano (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics 13*(3), 253–263.

Domowitz, I. and C. Hakkio (1985). Conditional variance and the risk premium in the foreign exchange market. *Journal of International Economics 19*(1), 47–66.

Duan, K. and A. Urquhart (2023). The instability of stablecoins. *Finance Research Letters 53*(103573).

Duffie, D., J. Pan, and K. Singleton (2000). Transform analysis and asset pricing for affine jump diffusions. *Econometrica 68*(6), 1343–1376.

Dumas, B., J. Flemming, and R. Whaley (1998). Implied volatility functions: Empirical tests. *Journal of Finance 53*(6), 2059–2106.

Dunis, C. and X. Huang (2002). Forecasting and trading currency volatility: An application of recurrent neural regression and model combination. *Journal of Forecasting 21*(5), 317–354.

Dupire, B. (1993). Model art. *Risk Magazine*, 118–120.

Dupire, B. (1994). Pricing with a Smile. *Risk Magazine*, 18–20.

Egloff, D., M. Leippold, and L. Wu (2010). The term structure of variance swap rates and optimal variance swap investments. *Journal of Financial and Quantitative Analysis 45*(5), 1279–1310.

Elman, J. (1990). Finding structure in time. *Cognitive Science 14*, 179–211.

Eross, A., F. McGroarty, A. Urquhart, and S. Wolfe (2019). The intraday dynamics of bitcoin. *Research in International Business and Finance 49*, 71–81.

Fengler, M. (2009). Arbitrage-free smoothing of the implied volatility surface. *Quantitative Finance 9*(4), 417–428.

Fengler, M. and W. Härdle (2003). The dynamics of implied volatilites: a common principal components approach. *Review of Derivatives Research 6*, 179–202.

Fengler, M., W. Härdle, and E. Mammen (2007). A semiparametric factor model for implied volatility surface dynamics. *Journal of Financial Econometrics 5*, 189–218.

François, P. and L. Stentoft (2021). Smile-implied hedging with volatility risk. *Journal of Futures Markets 41*(8), 1220–1240.

Fritsch, F. and J. Butland (1984). A method for constructing local monotone piecewise cubic interpolants. *SIAM Journal on Scientific and Statistical Computing 5*(2), 300–304.

Gal, Y. and Z. Ghahramani (2016). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. *arXiv Paper*.

Galai, D. (1979). A proposal for indexes for traded call options. *Journal of Finance 34*(5), 1157–1172.

Gandal, N., J. Hamrick, T. Moore, and T. Oberman (2018). Price manipulation in the bitcoin ecosystem. *Journal of Monetary Economics 144*, 86–96.

Garman, M. B. and S. W. Kohlhagen (1983). Foreign currency option values. *Journal of International Money and Finance 2*(3), 231–237.

Gatheral, J. (2006). The volatility surface: A practitioner's guide. *Wiley Finance Series*, 135–136.

Gatheral, J. and A. Jacquier (2014). Arbitrage-free svi volatility surfaces. *Quantitative Finance 14*(1), 59–71.

Gkillas, K., P. Katsiampa, C. Konstantatos, and A. Tsagkanos (2022). Discontinuous movements and asymmetries in cryptocurrency markets. *European Journal of Finance*.

Glorot, X., A. Bordes, and Y. Bengio (2011). Deep sparse rectifier neural networks. *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics 15*, 315–323.

Golez, B. and J. Jackwerth (2012). Pinning in the S&P 500 futures. . *Journal of Financial Economics 106*, 566–585.

Goll, T. and L. Rüschendorf (2001). Minimax and minimal distance martingale measures and their relatioship to portfolio optimization. *Finance and Stochastics 5*(4), 557–581.

Gonçalves, S. and M. Guidolin (2006). Predictable dynamics in the S&P 500 index options implied volatility surface. *Journal of Business 79*(3), 1591–1635.

Goodfellow, I., Y. Bengio, and A. Courville (2016). *Deeo Learning*. The MIT Press.

Grabbe, J. (1983). The pricing of call and put options on foreign exchange. *Journal of International Money and Finance 2*, 239–253.

Granger, C. and P. Newbold (1974). Spurious regressions in econometrics. *Journal of Econometrics 2*(2), 111–120.

Graves, A. (2012). *Supervised Sequence Labelling with Recurrent Neural Networks*. Springer Berlin, Heidelberg.

Griffin, J. and A. Shams (2018). Manipulation in the vix? *Review of Financial Studies 31*(4), 1377–1417.

Gruber, P. H., C. Trebaldi, and F. Trojani (2020). The price of the smile and variance risk premia. *Management Science 67*(7), 4056–4074.

Guyon, J. and J. Lekeufack (2023). Volatility is (mostly) path-dependent. *SSRN Working Paper*.

Haber, S. and W. Stornetta (1991). How to time-stamp a digital document. *Journal of Cryptology 3*(2), 99–111.

Haber, S. and W. Stornetta (1997). Secure names for bit-strings. *Proceedings of the ACM Conference on Computer and Communications Security*, 28–35.

Hagan, P., D. Kumar, A. Lesniewski, and D. Woodward (2002). Managing smile risk. *The Best of Wilmott 1*, 249–296.

Hamid, S. and A. Habib (2014). Financial forecasting with neural networks. *Academy of Accounting and Financial Studies Journal 18*(4), 37–55.

Harvey, C. and R. Whaley (1992). Market volatility prediction and the efficiency ofthe s&p 100 index options market. *Journal of Financial Economics 31*, 43–73.

Hastie, T., R. Tibshirani, and J. Friedman (2009). *The Elements of Statistical Learning*. Springer Series in Statistics.

Haugen, R., E. Talmor, and W. Torous (1991). The effect of volatility changes on the level of stock prices and subsequent expected returns. *The Journal of Finance 46*(3), 985–1007.

Hentschel, L. (2003). Errors in implied volatility estimation. *The Journal of Financial and Quantitative Analysis 38*(4), 779–810.

Heston, S. (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.

Heston, S. and S. Nandi (2000). A closed-form garch option valuation model. *Review of Financial Studies 13*, 585–625.

Hilliard, J. and J. Ngo (2022). Bitcoin: jumps, convenience yields, and option prices. *Quantitative Finance 22*(11), 2079–2091.

Hoang, L. and D. Baur (2021). How stable are stablecoins? *European Journal of Finance*.

Hoang, L. T. and D. G. Baur (2020). Forecasting bitcoin volatility: Evidence from the options market. *Journal of Futures Markets 40*(10), 1584–1602.

Hochreiter, S. and J. Schmidhuber (1997). Long short-term memory. *Neural Computaion 9*(8), 1735–1780.

Horvath, B., A. Muguruza, and M. Tomas (2021). Deep learning volatility: a deep neural network perspective on pricing and calibration in (rough) volatility models. *Quantitative Finance 21*(1), 11–27.

Hoster, J., S. Djurdjevic, H. Nguyen, and R. Slater (2018). Improving vix futures forecasts using machine learning forecasts using machine learning methods. *SMU Data Science Review 1*(4), Art. 6.

Hou, A., W. Wang, C. Chen, and W. Härdle (2020). Pricing cryptocurrency options. *Journal of Financial Econometrics 18*(2), 250–279.

Hu, J., W. Juo, and W. Haerdle (2019). Risk of bitcoin market: Volatility, jumps, and forecasts. *Cornell University Online*.

Huang, J.-Z., J. Ni, and L. Xu (2022). Leverage effect in cryptocurrency markets. *Pacific Basin Finance Journal 73*.

Hull, J. and A. White (2017). Optimal Delta Hedging for Options. *Journal of Banking and Finance 17*, 180–190.

Hung, J.-C., H.-C. Liu, and J. Yang (2020). Improving the realized garch's volatility forecast for bitcoin with jump-robust estimators. *North American Journal of Economics and Finance 52*.

Jackwerth, J. (1999). Option-implied risk-neutral distributions and implied binomial trees. *Journal of Derivatives 7*(2), 66–82.

Jackwerth, J. and M. Rubinstein (1996). Recovering probability distributions from option prices. *Journal of Finance 51*(5), 1611–1631.

Jaeger, H. (2001). The "echo state" approach to analysing and training recurrent neural networks-with an erratum note. *German National Research Center for Information Technology GMD Technical Report 34*.

Jalan, A., R. Matkovskyy, and S. Aziz (2021). The bitcoin options market: A first look at pricing and risk. *Applied Economics 53*(17), 2026–2041.

Jalan, A., T. Matkovskyy, and A. Urquhart (2021). What effect did the introduction of bitcoin futures have on the bitcoin spot market? *European Journal of Finance 27*, 1251–1281.

Jamshidian, F. (1993). Option and futures evaluation with deterministic volatilities. *Mathematical Finance 3*(2), 149–159.

Jansen, S. (2020). *Machine Learning for algorithmic Trading*. Packt Publishing.

Jarrow, R., Y. Kchia, M. Larsson, and P. Protter (2013). Discretely sampled variance and volatility swaps versus their continuous approximations. *Finance and Stochastics 17*(2), 305–324.

Jarrow, R. A. and M. Larsson (2012). The meaning of market efficiency. *Mathematical Finance 22*(1), 1–30.

Jeanblanc, M., M. Chesney, and M. Yor (2009). *Mathematical Methods for Financial Markets*. Springer Finance.

Jiang, G. and R. Oomen (2008). Testing for jumps when asset prices are observed with noise - A "swap variance" approach. *Journal of Econometrics 144*(2), 352–370.

Jiang, G. and Y. Tian (2005). The model-free implied volatility and its information content. *Review of Financial Studies 18*(4), 1305–1342.

Jordan, M. (1990). Attractor dynamics and parallelism in a connectionist sequential machine. *Artificial neural networks. concept Learning. IEEE Computer Society*, 112–127.

Kallsen, J. and A. Shiryayev (2002). The cumulant process and Esscher's change of measure. *Finance and Stochastics 6*(2), 313–338.

Kalman, B. and S. Kwasny (1992). Why tanh: choosing a sigmoidal function. *JCNN International Joint Conference on Neural Networks*, 578–581.

Kang, N. and J. Kim (2019). An empirical analysis of bitcoin price jump risk. *Sustainability 11*(7).

Kani, I., E. Derman, and M. Kamal (1996). Trading and hedging local volatility. *Goldman Sachs Quantitative Strategies Research Notes*.

Katsiampa, P. (2019). An empirical investigation of volatility dynamics in the cryptocurrency market. *Research in International Business and Finance 50*, 322–335.

Kearney, F., M. Cummins, and F. Murphy (2018). Forecasting implied volatility in foreign exchange markets: a functional time series approach. *European Journal of Finance 24*(1), 1–18.

Kelleher, J. (2019). *Deep Learning*. The MIT Press.

Kim, A., S. Trimborn, and W. K. Härdle (2021). Vcrix — a volatility index for cryptocurrencies. *International Review of Financial Analysis 78*, 101915.

Kim, H., K. Park, J. Jeon, C. Song, J. Bae, Y. Kim, and M. Kang (2021). Candidate point selection using a self-attention mechanism for generating a smooth volatility surface under the sabr model. *Expert Systems with Applications 173*(114640).

Kim, H. and C. Won (2018). Forecasting the volatility of stock price index: A hybrid model integrating lstm with multiple garch-type models. *Expert Systems with Applications 103*, 25–37.

Kingma, D. and J. Ba (2015). Adam: A method for stochastic optimization. *arXic Working Paper*.

Konstantinidi, E. and G. Skiadopoulos (2016a). How does the market variance risk premium vary over time? Evidence from S&P 500 variance swap investment returns. *Journal of Banking and Finance 62*, 62–75.

Konstantinidi, E. and G. Skiadopoulos (2016b). How does the market variance risk premium vary over time? evidence from sp 500 variance swap investment returns. *Journal of Derivatives 62*(1), 62–75.

Köchling, G., P. Schmidtke, and P. Posch (2020). Volatility forecasting accuracy for bitcoin. *Economics Letters 191.*

Lee, R. (2001). Implied and local volatilities under stochastic volatility. *International Journal of Theoretical and Applied Finance 4*(1), 45–89.

Lee, R. (2004). The moment formula for implied volatility at extreme strikes. *Mathematical Finance 14*(3), 469–480.

Lee, S. and P. Mykland (2008). Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies 21*, 2535–2563.

Leibniz, G. (1676). Memoir using the chain rule (2010). *TMME 7(2)*, 321–332.

Leontsinis, S. and C. Alexander (2017). Arithmetic variance swaps. *Quantitative Finance 17*(4), 551–569.

Levy, E. (1992). Pricing european average rate currency options. *Journal of International Money and Finance 11*(5), 474–491.

Lewis, A. (2000). Option valuation under stochastic volatility. *Finance Press.*

Li, L., A. Arab, J. Liu, J. Liu, and Z. Han (2019). Bitcoin options pricing using lstm-based prediction model and blockchain statistics. pp. 67–74.

Lim, K.-G. and C. Ting (2013). The term structure of s&p 100 model-free volatilities. *Quantitative Finance 13*(7), 1041–1058.

Lucey, B. and E. Tully (2006). Seasonality, risk and return in daily comex gold and silver data 1982–2002. *Applied Financial Economics 16*(4), 319–333.

Lucic, V. (2022). Btc inverse call and the standard FX framework. *SSRN Wokring Paper.*

Lui, D., Y. Liang, L. Zhang, P. Lung, and R. Ullah (2021). Implied volatility forecast and option trading strategy. *International Review of Economics and Finance 71.*

Madan, D., P. Carr, and E. Chang (1998). The variance gamma process and option pricing. *Review if Finance 2*(1), 79–105.

Madan, D. and F. Milne (1994). Contingent claims valued and hedged by pricing and investing in a basis. *Mathematical Finance 4*, 223–245.

Madan, D., S. Reyners, and W. Schoutens (2019). Advanced model calibration on bitcoin options. *Digital Finance 1*, 117–137.

Malz, A. (2000). Vega risk and the smile. *RiskMetrics Working Paper No. 99-06.*

Martin, I. (2013). Simple variance swaps. *NBER Working Papers from National Bureau of Economic Research.*

Matic, J., N. Packham, and W. Härdle (2023). Hedging cryptocurrency options. *Review of Derivatives Research 26*(1), 91–133.

McNeil, A. and R. Frey (2000). Estimation of tail-related risk measures for heteroscedastic financial time series: An extreme value approach. *Journal of Empirical Finance 7*(3), 271–300.

Meddahi, N. and E. Renault (2004). Temporal aggregation of volatility models. *Journal of Econometrics 119*(1), 355–379.

Medvedev, A. and O. Scaillet (2007). Approximation and calibration of short-term implied volatilities under jump-diffusion stochastic volatility. *The Review of Financial Studies 20*(2), 427–459.

Medvedev, N. and Z. Wang (2022). Multistep forecast of the implied volatility surface using deep learning. *Journal of Futures Markets 42*(4), 645–667.

Mercurio, F. (2003). Pricing and static replication of FX quanto options. *Banca IMI Working Paper*.

Merton, R. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics 3*(1), 125–144.

Mikkiliä, O. anbd Kanniainen, J. (2023). Empirical deep hedging. *Quantitative Finance 23*(1), 111–122.

Monoyios, M. (2004). Performance of utility-based strategies for hedging basis risk. *Quantitative Finance 4*(3), 245–255.

Musiela, M. and M. Rutkowski (2005). *Martingale Methods in Financial Modelling*. Stochastic Modelling and Applied Probability. Springer.

Nair, V. and G. E. Hinton (2010). Rectified linear units improve restricted boltzmann machines. *Proceedings of the 27th International Conference on International Conference on Machine Learning 8*, 807–814.

Nakamoto, S. (2009). Bitcoin: A peer-to-peer electronic cash system. *https://bitcoin.org/bitcoin.pdf*.

Nastasi, E., A. Pallavicini, and G. Sartorelli (2020). Smile modeling in commodity markets. *International Journal of Theoretical and Applied Finance 23*(3).

Neuberger, A. (1994). The log contract. *Journal of Portfolio Management 20*(2), 74–80.

Neuberger, A. (2012). Realized skewness. *Review of Financial Studies 25*(11), 3423–3455.

Neumann, M. and G. Skiadopoulos (2013). Predictable dynamics in higher-order risk-neutral moments: Evidence from the S&P 500 options. *Journal of Financial and Quantitive Analysis 48*(3), 947–977.

Newey, W. and K. West (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies 61*(4), 631–653.

Ni, S., N. Pearson, A. Poteshman, and J. White (2021). Does option trading have a pervasive impact on underlying stock prices. *The Review of Financial Studies 34*(4), 1952–1986.

Pagnottoni, P. (2019). Neural network models for bitcoin option pricing. *Frontiers in Artificial Intelligence 2*.

Pena, I., G. Rubio, and G. Serna (1999). Why do we smile? on the determinants of the implied volatility function. *Journal of Banking and Finance 23*(8), 1151–1179.

Platanakis, E. and A. Urquhart (2020). Should investors include bitcoin in their portfolios? a portfolio theory approach. *British Accounting Review 52*(4), 100837.

Prokopczuk, M., L. Symeonidis, and C. Wese Simen (2017a). Variance risk in commodity markets. *Journal of Banking and Finance 81*, 136–149.

Prokopczuk, M., L. Symeonidis, and C. Wese Simen (2017b). Variance risk in commodity markets. *Journal of Banking and Finance 81*, 136–149.

Qian, L., J. Wang, F. Ma, and Z. Li (2022). Bitcoin volatility predictability–the role of jumps and regimes. *Finance Research Letters 47*.

Reiswich, D. and U. Wystup (2010). A guide to FX options quoting conventions. *Journal of Derivatives 18*(2), 58–68.

Ritter, H. and T. Kohonen (1989). Self-organizing semantic maps. *Biological Cybernetic 61*, 241–254.

Robinson, A. and F. Fallside (1987). The utility driven dynamic error propagation network. *Technical Report CUED/F-INFENG/TR.1, Cambridge University Engineering Department*.

Rompolis, L. and E. Tzavalis (2017). Pricing and hedging contingent claims using variance and higher order moment swaps. *Journal of Quantitative Finance 17*(4), 531–550.

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review 65*(6), 386–408.

Rouge, R. and N. El Karoui (2000). Pricing via utility maximization and entropy. *Mathematical Finance 10*(2), 259–276.

Rubinstein, M. (1994). Implied binomial trees. *Journal of Finance 49*, 771–818.

Rumelhart, D., G. Hinton, and R. Williams (1986). Learning representations by back-propagating errors. *Nature 323*, 533–536.

Scaillet, O., A. Treccani, and C. Trevisan (2020). High-frequency jump analysis of the bitcoin market. *Journal of Financial Econometrics 18*(2), 209–232.

Schachermayer, W. (2002). Optimal investment in incomplete mfinancial markets. *Mathematical Finance Bachelier COngress*(2000), 427–642.

Schmidt, R. (2019). Recurrent neural networks (rnns): A gentle introduction and overview. *arXiv*.

Schneider, P. and F. Trojani (2018). (Almost) model-free recovery. *Journal of Finance 74*(1), 323–370.

Sepp, A. and P. Rakhmonov (2022). Log-normal stochastic volatility model with quadratic drift: Applications to assets with positive return-volatility correlation and to inverse martingale measures. *SSRN Working Paper*.

Sharpe, W. F. (1994). The sharpe ratio. *The Journal of Portfolio Management 21*(1), 49–58.

Shen, D., A. Urquhart, and P. Wang (2020). Forecasting the volatility of bitcoin: The importance of jumps and structural breaks. *European Financial Management 26*(5), 1294–1323.

Shen, D., A. Urquhart, and P. Wang (2022). Bitcoin intraday time series momentum. *Financial Reviews 57*(2), 243–251.

Shimko, D. (1993). Bounds of probability. *Risk 6*, 33–37.

Siu, T. K. and R. Elliott (2021). Bitcoin option pricing with a SETRA-GARCH model. *The European Journal of Finance 27*(6), 564–595.

Skiadopoulos, G., S. Hodges, and L. Clewlow (1999). The dynamics of the S&P 500 implied volatility surface. *Review of Derivatives Research 3*, 263–282.

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

Srivastava, N., G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research 15*(56), 1929–1958.

Sutskever, I., J. Martens, G. Dahl, and G. Hinton (2013). On the importance of initialization and momentum in deep learning. *Proceedings of the 30th International Conference on Machine Learning 28*(3), 1139–1147.

Teng, H. and W. Härdle (2022). Financial analytics of inverse BTC options in a stochastic volatility world. *SSRN Working Paper*.

Todorov, V. (2010). Variance risk-premium dynamics: The role of jumps. *Review of Financial Studies 23*(1), 345–383.

Tompkins, R. and R. D'Ecclesia (2006). Unconditional return disturbances: A non-parametric simulation approach. *Journal of Banking and Finance 30*, 287–314.

Trolle, A. and E. Schwartz (2010). Variance risk premia in energy commodities. *Journal of Derivatives 17*(3), 15–32.

Urquhart, A. (2016). The inefficiency of bitcoin. *Economic Letters 148*, 80–82.

Vähämaa, S. (2004). Delta hedging with the smile. *Financial Markets and Portfolio Management 18*(3), 241–255.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, and I. Poloshukhin (2017). Attention is all you need. *arXiv Working Paper*.

Venter, P. J. and E. Mare (2020). Garch generated volatility indices of bitcoin and crix. *Journal of Risk Financial Management 13*(6), 121.

Vrontos, S., J. Galakis, and I. Vrontos (2021). Implied volatility directional forecasting: a machine learning approach. *Quantitative Finance 21*(10), 1687–1706.

Wang, J.-N., H.-C. Liu, and Y.-T. Hsu (2020). Time-of-day periodicities of trading volume and volatility in bitcoin exchange: Does the stock market matter? *Finance Research Letters 34*(101243), 1–8.

Werbos, P. (1990). Backpropagation through time: What it does and how to do it. *Proceedings of the IEEE 78*(10), 1550–1560.

Whaley, E. (2000). The investor fear gauge. *The Journal of Portfolio Management 26*(3), 12–17.

Woebbeking, F. (2021). Cryptocurrency volatility markets. *Digital Finance 3*, 273–298.

Wolpert, D. H. and W. Macready (1995). No free lunch theorems for search. *SFI Working Paper*, 1–33.

Zhang, A., Z. C. Lipton, M. Li, and A. J. Smola (2023). *Dive into Deep Learning*. Cambridge University Press.

Zhang, Y., M. He, D. Wen, and Y. Wang (2022). Forecasting bitcoin volatility: A new insight from the threshold regression model. *Journal of Forecasting 41*(3), 633–652.

Zulfiqar, N. and S. Gulzar (2021). Implied volatility estimation of bitcoin options and the stylized facts of option pricing. *Financial Innovation 7*(1).

Appendices

APPENDIX A

CRYPTO OPTION PRICING

A.1 Derivation of the Greeks

We present a step-by-step derivation of the quanto inverse Greeks. For the sake of readability, we denote $S^{\$} = S$. First, we derive some basic formulae:

$$d_2 = \frac{\ln\left(\frac{S}{K}\right) + \left(r - \frac{1}{2}\sigma^2\right)\tau}{\sigma\sqrt{\tau}}, \qquad d_3 = \frac{\ln\left(\frac{S}{K}\right) + \left(r - \frac{3}{2}\sigma^2\right)\tau}{\sigma\sqrt{\tau}} = d_2 - \sigma\sqrt{\tau} \qquad (A.1)$$

Note that the standard normal probability function $\phi(\cdot)$ is given as:

$$\phi(d_2) = \phi(-d_2) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_2^2}.$$

We use this to simplify:

$$\phi(d_3) = \phi\left(d_2 - \sigma\sqrt{\tau}\right) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(d_2 - \sigma\sqrt{\tau}\right)^2} = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(d_2^2 - 2d_2\sigma\sqrt{\tau} + \sigma^2\tau\right)}$$

$$= \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}d_2^2}e^{d_2\sigma\sqrt{\tau}}e^{-\frac{1}{2}\sigma^2\tau} = \phi(d_2)e^{\ln\left(\frac{S}{K}\right) + \left(r - \frac{1}{2}\sigma^2\right)\tau}e^{-\frac{1}{2}\sigma^2\tau}$$

$$= \phi(d_2)\frac{S}{K}e^{(r-\sigma^2)\tau}. \qquad (A.2)$$

Next we show some basic derivatives:

$$\frac{\partial d_2}{\partial S} = \frac{\partial d_3}{\partial S} = \frac{1}{S\sigma\sqrt{\tau}}, \qquad (A.3)$$

as well as:

$$\frac{\partial d_2}{\partial \sigma} = \frac{-\ln\left(\frac{S}{K}\right) - r\tau - \frac{1}{2}\sigma^2\tau}{\sigma^2\sqrt{\tau}}, \qquad (A.4)$$

$$\frac{\partial d_3}{\partial \sigma} = \frac{-\ln\left(\frac{S}{K}\right) - r\tau - \frac{3}{2}\sigma^2\tau}{\sigma^2\sqrt{\tau}} = \frac{\partial d_2}{\partial \sigma} - \sqrt{\tau}, \qquad (A.5)$$

180

and lastly:

$$\frac{\partial d_2}{\partial \tau} = \frac{-0.5 \ln\left(\frac{S}{K}\right) + 0.5 r \tau - \frac{1}{4}\sigma^2 \tau}{\sigma \sqrt{\tau^3}} \tag{A.6}$$

$$\frac{\partial d_3}{\partial \tau} = \frac{-0.5 \ln\left(\frac{S}{K}\right) + 0.5 r \tau - \frac{3}{4}\sigma^2 \tau}{\sigma \sqrt{\tau^3}} = \frac{\partial d_2}{\partial \tau} - \frac{\sigma}{2\sqrt{\tau}}. \tag{A.7}$$

Using (A.1) and (A.3), we can now define the quanto inverse call delta, i.e. the sensitivity of the quanto inverse call price with respect to the underlying. We deliberately omit the quanto inverse put derivation.

$$\begin{aligned}
\delta := \frac{\partial f}{\partial S} &= \frac{\partial}{\partial S} e^{-r\tau} \left[\Phi(d_2) - e^{(\sigma^2 - r)\tau} S^{-1} K \Phi(d_3) \right] \\
&= e^{-r\tau} \left[\frac{\partial \Phi(d_2)}{\partial d_2} \frac{\partial d_2}{\partial S} - e^{(\sigma^2 - r)\tau} \left[-\frac{K}{S^2} \Phi(d_3) + \frac{K}{S} \frac{\partial \Phi(d_3)}{\partial d_3} \frac{\partial d_3}{\partial S} \right] \right] \\
&= e^{-r\tau} \left[\frac{\phi(d_2)}{S\sigma\sqrt{\tau}} - e^{(\sigma^2 - r)\tau} \left[-\frac{K}{S^2} \Phi(d_3) + \frac{K}{S^2 \sigma\sqrt{\tau}} \phi(d_3) \right] \right] \\
&= e^{-r\tau} \left[\frac{\phi(d_2)}{S\sigma\sqrt{\tau}} + e^{(\sigma^2 - r)\tau} \frac{K}{S^2} \Phi(d_3) - e^{(\sigma^2 - r)\tau} \frac{K}{S^2 \sigma\sqrt{\tau}} \phi(d_2 - \sigma\sqrt{\tau}) \right] \\
&= e^{-r\tau} \left[\frac{\phi(d_2)}{S\sigma\sqrt{\tau}} + e^{(\sigma^2 - r)\tau} \frac{K}{S^2} \Phi(d_3) - e^{(\sigma^2 - r)\tau} \frac{K}{S^2 \sigma\sqrt{\tau}} \phi(d_2) \frac{S}{K} e^{-(\sigma^2 - r)\tau} \right] \\
&= e^{(\sigma^2 - 2r)\tau} S^{-2} K \Phi(d_3).
\end{aligned}$$

In similar spirit, we can derive the quanto inverse call gamma, i.e. the second derivative of the quanto inverse call price with respect to the underlying:

$$\gamma := \frac{\partial \delta}{\partial S} = \frac{\partial}{\partial S} \left[e^{(\sigma^2 - 2r)\tau} S^{-2} K \Phi(d_3) \right] = e^{(\sigma^2 - 2r)\tau} S^{-3} K \left[\frac{\phi(d_3)}{\sigma\sqrt{\tau}} - 2\Phi(d_3) \right].$$

Next, we consider the quanto inverse call vega (ν). Using (A.1) and (A.6), we show:

$$\begin{aligned}
\nu := \frac{\partial f}{\partial \sigma} &= \frac{\partial}{\partial \sigma} e^{-r\tau} \left[\Phi(d_2) - e^{(\sigma^2 - r)\tau} S_t^{-1} K \Phi(d_3) \right] \\
&= e^{-r\tau} \left[\frac{\partial \Phi(d_2)}{\partial d_2} \frac{\partial d_2}{\partial \sigma} - \frac{K}{S} \left[\frac{\partial e^{(\sigma^2 - r)\tau}}{\partial \sigma} \Phi(d_3) + e^{(\sigma^2 - r)\tau} \frac{\partial \Phi(d_3)}{\partial d_3} \frac{\partial d_3}{\partial \sigma} \right] \right] \\
&= e^{-r\tau} \left[\phi(d_2) \frac{\partial d_2}{\partial \sigma} - e^{(\sigma^2 - r)\tau} \frac{K}{S} \left[2\tau\sigma\Phi(d_3) + \phi(d_2 - \sigma\sqrt{\tau}) \left(\frac{\partial d_2}{\partial \sigma} - \sqrt{\tau} \right) \right] \right] \\
&= e^{-r\tau} \left[\phi(d_2) \frac{\partial d_2}{\partial \sigma} - e^{(\sigma^2 - r)\tau} \frac{K}{S} \left[2\tau\sigma\Phi(d_3) + \phi(d_2) e^{d_2 \sigma\sqrt{\tau}} e^{-\frac{\sigma^2 \tau}{2}} \left(\frac{\partial d_2}{\partial \sigma} - \sqrt{\tau} \right) \right] \right] \\
&= e^{-r\tau} \left[\phi(d_2) \frac{\partial d_2}{\partial \sigma} - e^{(\sigma^2 - r)\tau} \frac{K}{S} \left[2\tau\sigma\Phi(d_3) + \phi(d_2) \frac{S}{K} e^{-(\sigma^2 - r)\tau} \left(\frac{\partial d_2}{\partial \sigma} - \sqrt{\tau} \right) \right] \right] \\
&= e^{-r\tau} \left[\phi(d_2) \sqrt{\tau} - 2 e^{(\sigma^2 - r)\tau} S^{-1} K \tau\sigma\Phi(d_3) \right].
\end{aligned}$$

Lastly, we present the time sensitivity theta (θ:

$$\vartheta := \frac{\partial f}{\partial \tau} = \frac{\partial}{\partial \tau} e^{-r\tau} \Phi(d_2) - e^{(\sigma^2 - 2r)\tau} S_t^{-1} K \Phi(d_3)$$

$$= \left[\frac{\partial e^{-r\tau}}{\partial \tau} \Phi(d_2) + e^{-r\tau} \frac{\partial \Phi(d_2)}{\partial d_2} \frac{\partial d_2}{\partial \tau} \right] - \frac{K}{S} \left[\frac{\partial e^{(\sigma^2 - 2r)\tau}}{\partial \tau} \Phi(d_3) + e^{(\sigma^2 - r)\tau} \frac{\partial \Phi(d_3)}{\partial d_3} \frac{\partial d_3}{\partial \tau} \right]$$

$$= \left[-re^{-r\tau} \Phi(d_2) + e^{-r\tau} \phi(d_2) \frac{\partial d_2}{\partial \tau} \right] -$$
$$\frac{K}{S} \left[\left(\sigma^2 - 2r\right) \Phi(d_3) e^{(\sigma^2 - 2r)\tau} + e^{(\sigma^2 - 2r)\tau} \phi(d_3) \frac{\partial d_3}{\partial \tau} \right]$$

$$= \left[-re^{-r\tau} \Phi(d_2) + e^{-r\tau} \phi(d_2) \frac{\partial d_2}{\partial \tau} \right] -$$
$$\frac{K}{S} \left[\left(\sigma^2 - 2r\right) \Phi(d_3) e^{(\sigma^2 - 2r)\tau} + e^{(\sigma^2 - 2r)\tau} \phi(d_2 - \sqrt{\tau}) \left(\frac{\partial d_2}{\partial \tau} - \frac{\sigma}{2\sqrt{\tau}} \right) \right]$$

$$= \left[-re^{-r\tau} \Phi(d_2) + e^{-r\tau} \phi(d_2) \frac{\partial d_2}{\partial \tau} \right] -$$
$$\frac{K}{S} \left[\left(\sigma^2 - 2r\right) \Phi(d_3) e^{(\sigma^2 - 2r)\tau} + e^{(\sigma^2 - 2r)\tau} \phi(d_2) \frac{S_T^\$}{K} e^{(r-\sigma^2)\tau} \left(\frac{\partial d_2}{\partial \tau} - \frac{\sigma}{2\sqrt{\tau}} \right) \right]$$

$$= \left[-re^{-r\tau} \Phi(d_2) + e^{-r\tau} \phi(d_2) \frac{\partial d_2}{\partial \tau} \right] - \left(\sigma^2 - 2r\right) \Phi(d_3) \frac{K}{S} e^{(\sigma^2 - 2r)\tau} -$$
$$e^{-r\tau} \phi(d_2) \left(\frac{\partial d_2}{\partial \tau} - \frac{\sigma}{2\sqrt{\tau}} \right)$$

$$= -re^{-r\tau} \Phi(d_2) + e^{-r\tau} \phi(d_2) \frac{\partial d_2}{\partial \tau} - \left(\sigma^2 - 2r\right) \Phi(d_3) \frac{K}{S} e^{(\sigma^2 - 2r)\tau} -$$
$$e^{-r\tau} \phi(d_2) \frac{\partial d_2}{\partial \tau} + e^{-r\tau} \phi(d_2) \frac{\sigma}{2\sqrt{\tau}}$$

$$= -re^{-r\tau} \Phi(d_2) - e^{(\sigma^2 - 2r)\tau} \left(\sigma^2 - 2r\right) \Phi(d_3) \frac{K}{S} + e^{-r\tau} \phi(d_2) \frac{\sigma}{2\sqrt{\tau}}.$$

We deliberately exclude the derivation of the other Greeks as well as those for put options for the sake of clarity and space. However, considering (A.1), (A.2), (A.3), (A.4) and (A.5) it is a rather trivial but gradual task to derive and present the missing Greeks.

APPENDIX B

THE LONG SHORT TERM MEMORY MODEL

Motivation: The significance of employing machine learning algorithms for financial data analysis cannot be understated. Recent technological advancements and the availability of large-scale datasets catalysed a resurgence of machine learning models. Although the application of these models to financial data is yet at its infancy, a limited number of research studies demonstrate promising results in the realms of pricing, hedging, and forecasting derivative data. Financial powerhouses like Bloomberg or Citadel committed substantial resources to explore and exploit this field. Although the foundations of these models often consist of straightforward linear algebraic equations, they still appear like a black box to numerous researchers.

Summary: We present the fundamental architecture of both feedforward and recurrent neural networks, explaining the mechanisms through which these models operate, learn, and generate predictions. We provide a comprehensive overview of the essential components constituting these models, alongside a rationale for employing neural networks in addressing predictive challenges. Ideas, derivation and information in this chapter are not my own. As a textbook chapter, all notations, definitions, figures, formulae and examples are inspired by and adopted from Goodfellow et al. (2016), Kelleher (2019), Schmidt (2019) Jansen (2020), Brownlee (2021) and Zhang et al. (2023).

B.1   Promise of Machine Learning

There has been a resurgence of interest in the application of self-learning algorithms to solve a broad range of problems in recent years.[1] The concept of machine learning (ML) has its origins in the 1960s, yet its core idea remained the same. Within this field, models autonomously "learn" and accumulate "experience" throughout a task, rather than relying on external sources for updates. The objective is to identify relationships, highlight key features, or generate entirely new elements in a self-learning fashion. Machine learning can be categorized into three primary paradigms: supervised learning, unsupervised learning, and reinforcement learning (RL). Supervised learning requires labelled data, wherein the model is trained on input data with known target outputs. Examples of supervised learning include the recognition of handwritten digits or the analysis of time series data. In contrast, unsupervised learning models lack access to target output data; instead, the goal is to reveal the structure of the input data. The probably most prominent model in this class is a Principal Component Analysis (PCA). Lastly, reinforcement learning does not require labeled data but strives to discover an optimal strategy within an Markovian environment by receiving rewards or penalties based on the model's actions. The ultimate goal is to maximize the reward over time, with self-driving cars serving as a notable application in this field. Particularly, the supervised learning model are of interest for our study as this subgroup can handle time series and generate predictions. Hence, we remove unsupervised and reinforcement learning from further consideration.

Among the supervised learning models, artificial neural networks (ANN) are perhaps the most renowned. These models attempt to mimic the structure and information processing mechanisms of the human brain. Similar to a human neuron, information is processed sequentially through action (input), perception (activation), and reaction (output). Models in which information is processed unidirectionally are referred to as feedforward neural networks (FNN), whereas those with recurrent or feedback loops constitute the class of recurrent neural networks (RNN). The earliest implementations of FNN can be traced back to the 1950s and have since spawned a variety of distinct models. Nonetheless, the most prevalent network among practitioners remains the Multilayer Perceptron (MLP), as introduced by Rumelhart et al. (1986).[2] Recurrent neural networks represent an extension of ANN which incorporate an internal memory function that retains information from previous inputs. Consequently, these models are better suited for handling sequential input data, such as translation, speech recognition, or time series analysis. Numerous variations of RNNs exist, each addressing specific limitations of the basic model.[3]

Recent technologically advancements in availability of large-scale data make machine learning models an attractive alternative to traditional statistical models. Both academics and industry professionals increasingly adopt ML models in financial data analysis due to their capacity to process vast amounts of data, identify hidden patterns, and make data-driven decisions. Common areas of applications include, among others, hedging single options (Mikkiliä, 2023) or option portfolios (Buehler et al., 2019), pricing (Hoang and Baur, 2021) derivatives or forecasting Vrontos et al. (2021) volatility as discussed in the literature review. The ML models' ability to capture non-linear, regime-specific dynamics within underlying data sets,

---

[1]The term "algorithm" defines a static code protocol a computer follows step-wise till it reaches its end.

[2]See a simple perceptron model (Rosenblatt (1958)), the radial basis function networks (Broomhead and Lowe (1988)) or Ritter-Kohonen maps (Ritter and Kohonen (1989)) to name a few

[3]See Elman networks (Elman (1990)), Jordan network (Jordan (1990)) or echo state network (Jaeger (2001)). However, discussing these in detail would exceed the scope of this work.

as well as their minimal reliance on data assumptions, position these models as the most promising innovations in contemporary financial research.

In the following, Section B.2 describes the basic architecture of a feedforward neural network and its shortcomings for sequential data; and Section B.3 introduces the recurrent extension of a FFN and emphasises the use of a long-short term memory model.

B.2 Feedforward Neural Network

The multilayer perceptron serves as the plain vanilla model for supervised feedforward neural networks, providing a base architecture upon which other, more sophisticated models are built. Consequently, we explain the essential architecture and procedure in greater depth. A MLP is a function $f_{\boldsymbol{\omega}} : \mathcal{X} \to \mathcal{Y}$, mapping the d-dimensional instance domain, sometimes referred to as training or input set, $\mathcal{X} \subset \mathbb{R}^d$ to the labelled domain $\mathcal{Y} \subset \mathbb{R}^{d_o}$, given the internal state of weights $\boldsymbol{\omega}$. This function processes labeled input data, or samples, $\boldsymbol{x} \in \mathbb{R}^d$, in form of a d-dimensional vector and maps it to a d_o-dimensional output vector $\hat{\boldsymbol{y}} \in \mathbb{R}^{d_o}$, attempting to estimate a conditioned probability distribution $P(\boldsymbol{y}|\boldsymbol{x})$. One or more hidden layers $\boldsymbol{h}^{(l)}$, where $l \in \{1, ..., L\}$ addresses a specific layer, bridge these two vectors. The number of neurons, or hidden units $|\boldsymbol{h}^{(l)}| = N_l$ in $\boldsymbol{h}^{(l)} \in \mathbb{R}^{1 \times N_l}$ may vary. In a sense, the input and output vectors can be regarded as layers themselves, albeit not hidden. A deep neural network describes a network with $L \geq 1$. We adhere to the standard literature notation where a bracketed superscript denotes the hidden layer, bold letters represent matrices, and subscripts address specific elements within a matrix.

B.2.1 Forward Propagation

The hidden layers are the processing units and form the backbone of any MLP. Consider the simple case of one hidden layer, i.e $L = 1$. Information flows from the input vector to the neurons, gets processed, and passed forward to the output vector. These flows are unidirectional, and we define the output of the i-th neuron of the hidden layer as $a_i^{(1)}$. In a fully connected MLP, each neuron is connected to all elements of the input vector, with individual coefficients, or weights, associated with these connections. We denote the weight from the i-th input element to the j−th neuron in first layer as $w_{(i,j)}^{(1)}$. In general, MLPs do not impose any constraints on these weights, but these range typically between 1 and 0 or -1, depending on the problem and modern libraries like `Keras` or `Scitlearn` allow to set borders. Nonetheless, these limits can profoundly impact the learning structure of the model, and theoretically, there are no absolute restrictions, as some MLPs have weights exceeding one, see this discussion for an example.

Each neuron tries to identify a specific pattern or key characteristic within the input data. This "divide and conquer" approach (Kelleher, 2019) combines the identified components and tries to capture the bigger picture, similar to a mosaic. As such, the selection of the number of layers and neurons is critical, as too many may result in capturing redundant information, while too few could neglect essential content. The operating principle of a neuron j is as simple as effective: First, it takes multiple inputs, which are the outputs of previous neurons, or the input vector elements, and multiplies them with their associated weight and sums them up to a value $z_j^{(1)}$. Additionally, it adds a bias term $b_j^{(1)}$ to the sum in order to parallel shift the weighted sum and consequently gaining more flexibility for the model fitting.[4] It is common

---

[4] The bias term is an additional parameter added to the weighted sum of the inputs. It serves as an offset,

to replace the bias by an additional weight $w_j^{(1)}$ and we will do so for further calculations. In the second step, the sum is passed through an activation function $\sigma^{(l)}$ at each layer l which produces in the neuron's output:

$$z_j^{(1)} = w_{(1,j)}^{(1)} x_1 + w_{(2,j)}^{(1)} x_2 + \dots + w_{(d,j)}^{(1)} x_d - w_j^{(1)}$$
$$a_j^{(1)} = \sigma_j^{(1)} \left(z_j^{(1)} \right).$$

The final model output follows an identical procedure with updated values and weights:

$$z_j^{(o)} = w_{(1,j)}^{(o)} a_1^{(1)} + w_{(2,j)}^{(o)} a_2^{(1)} + \dots + w_{(N_1,j)}^{(o)} a_{N_1}^{(1)} - w_j^{(o)}$$
$$\hat{y}_j = \sigma_j^{(o)} \left(z_j^{(o)} \right),$$

where $w^{(o)}$ denotes the weighs from neuron to output vector and $\sigma^{(o)}$ defines an (optional) activation function before the output. This formula can be expressed more general using the matrix notation:[5]

$$\boldsymbol{Z}^{(1)} = \boldsymbol{W}^{(1)} \boldsymbol{x} + \boldsymbol{b}^{(1)}$$
$$\boldsymbol{h}^{(1)} = \sigma^{(1)} \left(\boldsymbol{Z}^{(1)} \right)$$
$$\boldsymbol{Z}^{(o)} = \boldsymbol{W}^{(o)} \boldsymbol{h}^{(1)} + \boldsymbol{b}^{(o)}$$
$$\hat{\boldsymbol{y}} = \sigma^{(o)} \left(\boldsymbol{Z}^{(o)} \right)$$

with

$$\boldsymbol{x} = [x_1, x_2, \dots, x_d]^\mathsf{T}, \quad \boldsymbol{b}^{(1)} = \left[w_1^{(1)}, w_2^{(1)}, \dots, w_{N_1}^{(1)} \right]^\mathsf{T}, \quad \boldsymbol{b}^{(o)} = \left[w_1^{(o)}, w_2^{(o)}, \dots, w_{d_o}^{(o)} \right]^\mathsf{T}$$

and

$$\boldsymbol{W}^{(1)} = \begin{bmatrix} w_{(1,1)}^{(1)} & w_{(2,1)}^{(1)} & \cdots & w_{(d,1)}^{(1)} \\ w_{(1,2)}^{(1)} & w_{(2,2)}^{(1)} & \cdots & w_{(d,2)}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ w_{(1,N_1)}^{(1)} & w_{(2,N_1)}^{(1)} & \cdots & w_{(d,N_1)}^{(1)} \end{bmatrix}, \quad \boldsymbol{W}^{(o)} = \begin{bmatrix} w_{(1,1)}^{(o)} & w_{(2,1)}^{(o)} & \cdots & w_{(N_1,1)}^{(o)} \\ w_{(1,2)}^{(o)} & w_{(2,2)}^{(o)} & \cdots & w_{(N_1,2)}^{(o)} \\ \vdots & \vdots & \ddots & \vdots \\ w_{(1,d_o)}^{(o)} & w_{(2,d_o)}^{(o)} & \cdots & w_{(N_1,d_o)}^{(o)} \end{bmatrix}.$$

Activation functions serve a crucial role as they decide whether a neuron is activated or not. These functions must be differentiable and have non-linear properties to facilitate non-linear mapping functions. A straightforward contradiction-based argument explains why: Assume a linear activation function, i.e. the output of each neuron is simply the weighted sum of the previous. The whole networks output would be trivial and simply resemble a linear regression. The optimal activation function is typically identified through a trial-and-error

---

allowing the activation function to shift along the input axis, providing greater flexibility for the model fitting. In a sense this resembles the constant b for a linear function $y = ax + b$.

[5]Note that out initial choice of $L = 1$ can be generalised by stacking up hidden layers, e.g. $\boldsymbol{h}^{(2)} = \sigma_2^{(1)} \left(\boldsymbol{Z}^{(2)} \right)$ up to $\boldsymbol{h}^{(L)} = \sigma_L^{(L)} \left(\boldsymbol{Z}^{(L)} \right).$

process. Table B.2.1 presents a selection of prevalent activation functions, illustrating their mathematical representations, derivatives, and visual demonstrations. In the early stages of neural network development, threshold functions were the standard choice for activation functions, comprising simple binary indicator functions. With the evolution of gradient-based learning models, the sigmoid function emerged as a natural alternative, offering a smooth, differentiable approximation of the threshold function. Kalman and Kwasny (1992) advocate the use of the hyperbolic tangent (tanh) function which maps input values to an interval between -1 and 1. Both the sigmoid and tanh functions enjoy considerable popularity due to their smooth, differentiable characteristics; however, the tanh function exhibits point symmetry about the origin which may show advantages in modelling. The rectified linear unit (ReLU) has probably become the most popular and widely-used choice, due to its simplicity of implementation and exceptional performance (Nair and Hinton (2010), Glorot et al. (2011)).

Table B.2.1: Selection of Activation Functions

A selection of prevalent activation functions, illustrating their mathematical representations, derivatives, and visual demonstrations.
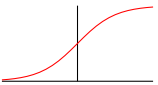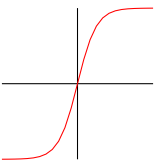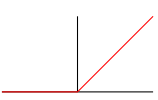
| Name	Function	Derivative	Figure
Threshold	$\sigma(x) = \mathbb{1}_{\{x > a\}}$	$\sigma'(x) = 0$	
Sigmoid	$\sigma(x) = \frac{1}{1+e^{-x}}$	$\sigma'(x) = \sigma(x)(1 - \sigma(x))^2$	
tanh	$\sigma(x) = \frac{1-e^{-2x}}{1+e^{-2x}}$	$\sigma'(x) = 1 - \sigma(x)^2$	
ReLU	$\sigma(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0. \end{cases}$	$\sigma'(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0 \end{cases}$	

Figure B.2.1 illustrates the standard MLP architecture, featuring one hidden layer with four neurons, a three-dimensional input vector, and a two-dimensional output.[6] The figure demonstrates the feedforward process: Each element of the input vector is multiplied by its corresponding weight, and the resulting sum (z) is forwarded to each neuron in the hidden layer. Subsequently, this value undergoes the activation function σ before being passed on to the output. $\hat{\boldsymbol{y}}$. At this stage, the feedforward process ends and the forecast is compared with

---

[6]This illustration follows recent literature conventions, where information flows from left to right. However, this choice is primarily based on personal preference and could alternatively depict information flowing from bottom to top, as was presented in the 1990s. This structure can be employed to address both classification and regression problems. The two-dimensional output might represent an action that needs to be taken, such as a buy or sell decision, or a judgment, such as black or white. The input for these two classification problems could consist of sentiment data or pixel values. It could also be used to predict the next two time steps of a time series. For this regression problem the input would like be the past three time steps or two time steps and an exogenous variable.

the pre-determined target output and prepared for next steps of the backpropagation; hence, the terminus of supervised learning with labelled data.

Figure B.2.1: Example Computational Graph of a Multilayer Perceptron

Simplistic visualisation of a fully connected MLP network with $L = 1$ hidden layer composed of four neurons, i.e. $N_1 = 4$. The input is a three dimensional vector $\boldsymbol{x}$ which results in a two-dimensional output vector $\hat{\boldsymbol{y}}$. Each neuron is connected to each and every element in the layers before and after, resulting 20 total weights in this system. We omit some weight descriptions for the sake of clarity and show the calculation within a neuron for only two neurons. The lower plot depicts an alternative visualisation of a MLP where the layer is summarised as a box. The filled arrow heads depict a matrix operation.
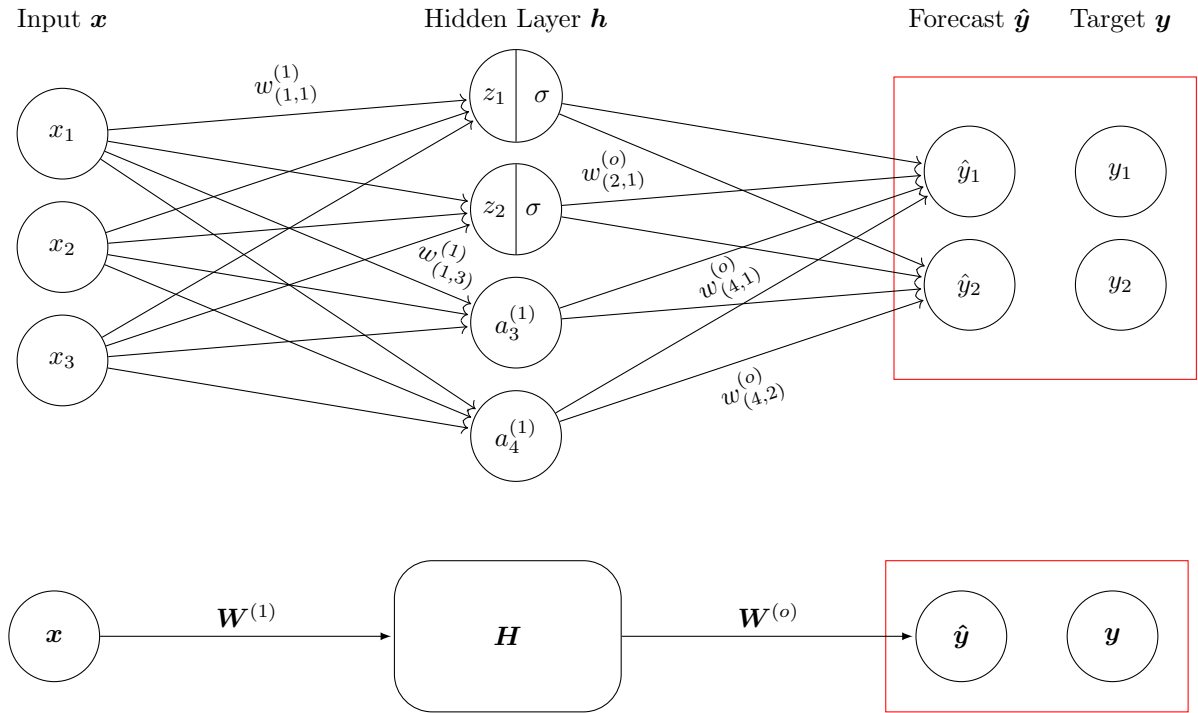
B.2.2 Backporpagation

For the sake of readability, let us keep the one-layer framework $L = 1$. A fully connected MLP is typically expressed in the form:

$$f(\boldsymbol{x}; \boldsymbol{\omega}),$$

where f denotes the mapping function, $\boldsymbol{x}$ reflects the input sample, and the internal state $\boldsymbol{\omega}$ is given as the combination of all available weights and biases within the network defines the model's internal state:

$$\boldsymbol{\omega} = \left(w_{(1,1)}^{(1)}, w_{(1,2)}^{(1)}, ..., w_{(N_1,N_{d_o})}^{(o)}, w_1^{(1)}, w_2^{(1)}, ...w_{d_o}^{(o)} \right).$$

Initially, the model selects $\boldsymbol{\omega}$ randomly, which in the majority of cases leads to suboptimal outputs. Consequently, the network must identify the optimal weights through a process known as learning. This procedure involves three steps: First, the model generates a forecast $\hat{\boldsymbol{y}}$ for a given sample $\boldsymbol{x}$, as discussed in (B.2.1). Second, the model compares the forecast with the actual value $\boldsymbol{y}$, which we highlight using the red square in Figure B.2.1, using a

pre-defined evaluation metric –or loss function– $\ell(\hat{\boldsymbol{y}}, \boldsymbol{y})$.[7] The choice of an appropriate loss function is often related to the activation function in the output layer, and there is a wide variety of potential options. Some of the most common loss functions include:

- **Mean Square Error (MSE):**
  This loss function is used mainly for regression problems and prediction of real numbers. The activation function in the last layer is ideally linear because a non-linear activation would skew the result. It is calculated as:

$$\ell(\hat{\boldsymbol{y}}, \boldsymbol{y}) = \frac{1}{d_o} \sum_{i=1}^{d_o} (y_i - \hat{y}_i)^2 .$$

  Further extensions of this function are Mean Square Logarithmic Error (MSLE) or the Mean Absolute Error (MAE) loss.

- **Logarithmic Loss (LS):**
  This approach is commonly used for classification problems. The ideal activation function is the sigmoid or tanh as these outputs might be interpreted as probabilities. This loss function is often referred as cross-entropy loss and is calculated as:

$$\ell(\hat{\boldsymbol{y}}, \boldsymbol{y}) = -\frac{1}{d_o} \sum_{i=1}^{d_o} (y_i \ln(\hat{y}_i) + (1 - y_i) \ln(1 - \hat{y}_i)) .$$

In the final step, the main objective of the learning process is to optimise $\boldsymbol{\omega}$, i.e. finding:

$$\underset{\boldsymbol{\omega}}{\arg \min} \ \ell(\hat{\boldsymbol{y}}, \boldsymbol{y}).$$

Learning, in essence, is a $|\omega|$-dimensional optimisation problem, which is addressed by a combination of the *gradient descent* and *backpropagation* algorithm. It is important to note that these terms should not be used interchangeably: the backpropagation algorithm efficiently computes the gradient, while the gradient descent refers to the optimization process. The gradient descent algorithm is relatively intuitive, as it uses the gradient $\nabla \ell$ to provide information on the slope and directional changes required to locate the minima of the surface. However, modern AI datasets contain vast amounts of data, which can quickly strain standard computational resources. Consequently, processing all samples through the network before updating the internal set of weights becomes a time-consuming task. To address this issue, there are three distinct gradient descent algorithms, which differ only in their batches B.[8] A batch is a parameter chosen arbitrarily to define the number of samples $\boldsymbol{x}$ that must pass through the entire network before updating the model weights. In fact, the current literature convention illustrates input samples not as single elements $\boldsymbol{x}$ but rather as a matrix/tensor $\boldsymbol{X}$ representing a batch of n samples. Based on this concept, the three primary gradient descent algorithms are:

- **Batch Gradient Descent:**
  This approach considers only one batch, i.e. the batch size is equal to the length of the

---

[7]Although the terms "loss" and "cost" are sometimes used interchangeably, there is a subtle distinction between these two: The loss function pertains to each sample, while the cost function represents the average value across multiple loss functions.

[8]The idea of approaching a minima using the gradient of a function is fairly nothing new. In fact, Cauchy (1847) is often referred to be the founder of this technique and cited frequently, see Goodfellow et al. (2016).

training set $\mathcal{X}$. Hence the cost function is calculated using the loss functions of every sample in the entire training dataset. It has the advantage of being a precise method to estimate the minimum but requires time and hardware resources since the output for every sample needs to be stored. The weight is updated after every sample has passed through the forward process, based on the gradient of the cost function.

- **Stochastic Gradient Descent:**
  The stochastic gradient descent (SGD) algorithm is an extension of the classic (batch) gradient descent algorithm. The main characteristic is the batch size of one, i.e. the internal state of the weights is updated based on the loss function after every randomly chosen sample that went through the network. The internal state is updated each sample. This makes the whole procedure rather fast and gives insight of the model performance but has a high probability of resulting a noisy learning curve which is prone to jumps. Note that it is highly unlikely to get the same results when running the data multiple times through the model because samples are selected randomly which influences the weight updates directly.

- **Mini-Batch Gradient Descent:**
  This method is a combination of the previous two and enjoys great popularity among practitioners. Batches in this approach are greater than one but smaller than the training set. The weights are updated based on the cost function over $n = |B|$ randomly selected samples.

The number of times all batches, i.e. the all samples in the dataset run once through the network is called **epoch** E. Here again, it is not guaranteed that the gradient will ultimately lead to the minima after one epoch. Hence, multiple iterations are needed to find the optimum weights. There is no magic formula to find the optimal number of iterations but plotting the error-function lead to the correct number.

In contrast to the analytical derivation, the numerical calculation of the gradient can be highly resource-intensive. The *backpropagation* algorithm, as introduced by Rumelhart et al. (1986), calculates the gradient in a cost-efficient and fast manner which is then forwarded to the gradient descent algorithm for the actual learning. Recall the general chain rule of calculus:

Theorem 1 (Chain Rule). *For $\boldsymbol{x} \in \mathbb{R}^d$, $\boldsymbol{y} \in \mathbb{R}^{d_o}$ and $g : \mathbb{R}^d \to \mathbb{R}^{d_o}, f : \mathbb{R}^{d_o} \to \mathbb{R}$ with $z = f(\boldsymbol{y})$ and $\boldsymbol{y} = g(\boldsymbol{x})$ it holds:*

$$\frac{\partial z}{\partial x_i} = \sum_{j=1}^{d_o} \frac{\partial z}{\partial y_j} \frac{\partial y_j}{\partial x_i},$$

where the subscript i denotes the $i-th$ element of the vector. More general, the chain rule states:

$$\nabla_{\boldsymbol{x}} z = \left(\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} \right)^\mathsf{T} \nabla_{\boldsymbol{y}} z,$$

where $= \left(\frac{\partial \boldsymbol{y}}{\partial \boldsymbol{x}} \right)$ is the Jacobian matrix of g and ∇ denotes the gradient vector.

Proof. See for example Leibniz (1676). □

Determining the partial derivative of each weight in the network is the primary objective of the backward propagation, specifically, computing $\nabla_{\boldsymbol{w}} \ell(\hat{\boldsymbol{y}}, \boldsymbol{y})$ efficiently. Referring back to

Figure B.2.1, it is obvious that the output vector $\hat{\boldsymbol{y}}$ has the greatest influence on the overall loss function. Computing the partial derivatives of $\ell\,(\hat{\boldsymbol{y}}, \boldsymbol{y})$ with respect to each element in $\hat{\boldsymbol{y}}$ is straightforward and solely depends on the activation function in the output layer. However, the forecast itself is merely a combination of outputs from previous layers. The weights to and from these preceding layers are not directly connected to the loss functions. Consequently, the algorithm proceeds backward to compute the partial derivatives of each weight. Following Graves (2012), and for the sake if clarity, it is helpful to introduce:

$$\delta_j = \frac{\partial \ell\,(\hat{\boldsymbol{y}}, \boldsymbol{y})}{\partial z_j},$$

with z_j being the summed weights of previous outputs. Moving from the output to the last hidden layer, we get:

$$\delta_i = \frac{\partial \ell\,(\hat{\boldsymbol{y}}, \boldsymbol{y})}{\partial a_i}\frac{\partial a_i}{\partial z_i} = \frac{\partial a_i}{\partial z_i}\sum_{k=1}^{|\hat{\boldsymbol{y}}|}\frac{\ell\,(\hat{\boldsymbol{y}}, \boldsymbol{y})}{\partial z_k}\frac{\partial z_k}{\partial a_i} \tag{B.1}$$

where $i \in \{1, 2, .., N_1\}$ addresses a node in the hidden layer, z_i is again the summed weight and a_i describes the output of a neuron in the layer. Note that the loss function depends on each node solely through the node's output. Taking the derivatives of these two yields:

$$\delta_i = \sigma'(a_i)\sum_{k=1}^{|\hat{\boldsymbol{y}}|}\delta_k w_{(i,k)}^{(o)}.$$

After all δ are calculated, we can use these to derive:

$$\frac{\partial \ell\,(\hat{\boldsymbol{y}}, \boldsymbol{y})}{\partial w_{(i,j)}} = \frac{\partial \ell\,(\hat{\boldsymbol{y}}, \boldsymbol{y})}{\partial a_j}\frac{\partial a_j}{\partial w_{i,j}} = \delta_j a_i$$

Or, more general:

$$\nabla_{\boldsymbol{\omega}}\ell\,(\hat{\boldsymbol{y}}, \boldsymbol{y}) = \left(\frac{\partial \boldsymbol{Z}}{\partial \boldsymbol{\omega}}\right)^{\mathsf{T}}\nabla_{\boldsymbol{Z}}\ell. \tag{B.2}$$

After each batch, the model then updates the internal weights:

$$\boldsymbol{\omega}_{new} = \boldsymbol{\omega}_{old} - (\eta\nabla_{\boldsymbol{\omega}}\ell\,(\hat{\boldsymbol{y}}, \boldsymbol{y})), \tag{B.3}$$

where η is an adjustable learning rate. Note that in this vector operation, each gradient is treated distinctly. The learning rate determines the extent to which the gradient (slope and direction) is used to update the weights. It plays a critical role in the timing and precision of the optimal weights. If the rate is too high, there is a possibility that the minimum is never detected but rather oscillates around it. If it is too low, the learning process might consume excessive resources.

B.2.3 Optimise the Optimisation

We presented the fundamental architecture and underlying mathematics of a plain vanilla FFN. This model can be employed to address a diverse array of problems. Specifically, we aim to explore its application in time series forecasting. In this setup, we would provide

the model a training set of p lagged data as input and target the step ahead prediction as output. An optimal ML predictive model would learn from the training data and extrapolate its acquired "knowledge" to a test set it has never encountered before. A popular metric to evaluate the goodness-of-fit is the loss function of the model, which illustrates how ℓ develops as the model trains over epochs. In an ideal scenario, the loss function would rapidly converge toward zero. The next step is to estimate the generalisation error, i.e. the difference between the loss function on the train and test set. In the optimal case, both loss functions would converge toward zero, indicating a flawless fit.

As model developer, we have great influence over the loss function's shape. Obvious approaches include adjusting hyperparameters such as the number of layers, nodes, and activation functions. However, Sutskever et al. (2013) emphasise that manipulating the input data itself likely has the most significant impact on the overall fit . Indeed, Wolpert and Macready (1995) highlight that learning algorithms perform better on data with certain distributions. In the following, we will introduce a selection of data manipulation techniques and architectural implementations aimed at enhancing overall model performance. Note that there exists a plethora of variations in the field of data manipulation and model enhancement. However, explaining each in greater detail would exceed the scope of this work.[9]

Scaling

Data scaling within the range $[a, b]$ presents a prominent preprocessing approach. This technique involves transforming the time such that the minimum value is assigned to a, the maximum corresponds to b, and the other fall within the interval. Typically, a is chosen to be either -1 or 0, while b is set 1. This proves beneficial, as most activation functions are particularly sensible within these intervals. More important, this method preserves the original distribution of the data. In general, using scaled data improves the overall model performance and allows for faster convergence. One reason is that the gradients with respect to different parameter will be more balanced if all features are on the same scale.

Early Stopping

Early stopping provides a conventional regularisation technique for deep neural networks. This method restricts number of training epochs and monitors the train (valuation) loss throughout the training process. It terminates the training process when the loss exhibits no improvement exceeding a predetermined threshold over a defined number of epochs. The main advantage is the time-saving aspect. In the context of large-scale models, this may yield considerable time and cost savings, which are of paramount importance in practical applications.

Checkpoint

The loss function does not decrease monotonously. In certain occasions, and depending on the batch size, the loss function may experience an abrupt increase. Hence, it is not assured that the overall best model fit is attained at the end of training, that is, after all epoch have been executed. The checkpoint method captures snapshots subsequent to each elapsed epoch. It stores the initial $\boldsymbol{\omega}$ internally and compares the loss after each epoch. Upon observing an enhancement in the loss function, it overwrites $\boldsymbol{\omega}$ with the current model weights. This process ensures selecting the best model, irrespective of the final update.

---

[9]For a detailed description of the following, as well as additional techniques, please see the original source of the definitions from Zhang et al. (2023).

Dropout

The Dropout technique (Srivastava et al., 2014) serves as the standard approach to avoid overfitting in ML modelling. As the name suggests, this method drops out neurons within each hidden layer at a predetermined probability. Figure B.2.2 is copied from Zhang et al. (2023) and demonstrates the mechanism in action. On the left, we have a fully connected MLP with one hidden layer compromised of five neurons. On the right side, we apply the Droput and see that the second and fifth neuron are deactivated during the forward process. As a result, the associated weights do not partake in the training process, thus reducing the likelihood of overfitting. As the computations do not rely on the two excluded neurons, their gradients vanish in the backpropagation algorithm. The Dropout technique is both computationally efficient and unrestricted in terms of model selection.

Figure B.2.2: Dropout Visualisation

Visual representation of a plain vanilla MLP network with one hidden layer before (left) and after (right) the Dropout function. This figure is copied from Zhang et al. (2023). All credits belong to the authors. For a detailed revision, please see Zhang et al. (2023) and their corresponding Book Website.
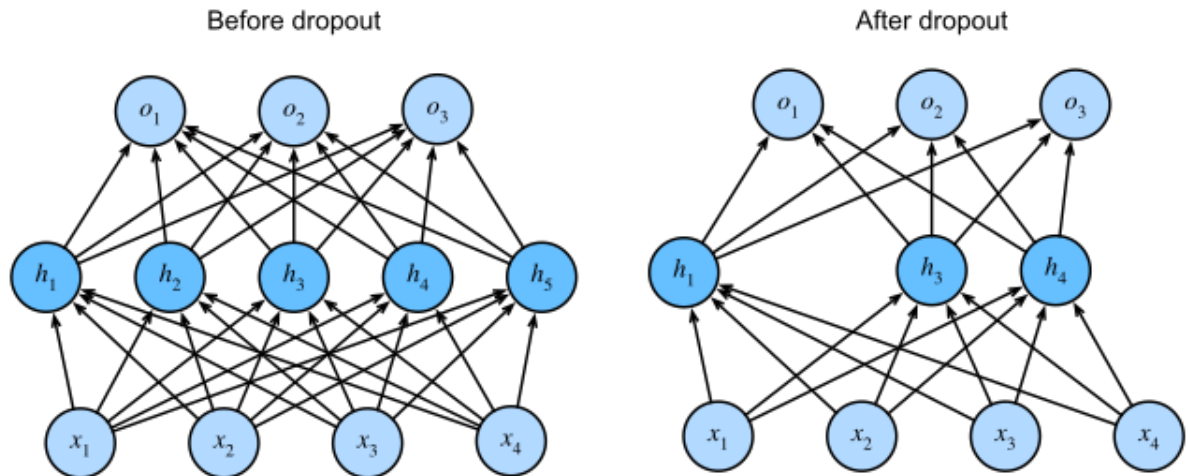
The Dropout function has a desirable side effect as it can be used as a Bayesian approximation (Gal and Ghahramani, 2016). The standard use of the dropout function is to disable connections during training randomly, but activate all connections for prediction. However, by actively keeping random connections disabled during prediction, each predictive result differs slightly for the same input. The simulation of multiple predictions then generates a histogram or density function of the possible predictions which allows us to estimate the prediction uncertainty and evaluate the model's reliability.

B.3   Recurrent Neural Network

The multi-layer perceptron model presents a reasonable attempt to predict (financial) time series data, despite its primary function being pattern recognition. It is important to note that the output relies solely on the arbitrarily selected (isolated) input from the training set. As such, it fails to capture the trend, regime or momentum prior to the input sample. It ignores favorable characteristics of the time series such as autocorrelation. To address this, we

introduce recurrent architectures, each exhibiting distinct properties that address particular challenges intrinsic to financial data.

Recurrent neural networks form an own class of neural networks that allow previous information to influence current operations. This characteristic makes these models particularly well-suited for sequential data, such as time series, and allows for the potential capture of different regimes. For this, the base architecture requires sequential input data in the form $\boldsymbol{X}_t \in \mathbb{R}^{n \times d}$, which is a mini-batch of sorted samples of n sequences. The current time step of variables will be denoted with a subscript. For readability, we keep the assumption of a single hidden layer and define the hidden state as $\boldsymbol{H}_t \in \mathbb{R}^{n \times N_1}$ at time t.[10] Note the time-dependency, i.e. the hidden state at a previous time will have direct influence on current calculations. As such, we need to store the hidden layer from operations and define a temporal dependency for the operations. These recurrent or hidden-to-hidden connections are captured by a weight matrix $\widetilde{\boldsymbol{W}} \in \mathbb{R}^{N_1 \times N_1}$ which connects hidden layers across time and is shared throughout the model.[11] The purpose of this additional parameter is to be an internal memory, which, unlike a simple FNN, uses information from possibly the entire history to map an output. The key components are iterative loops across time steps. More specifically, since neurons attempt to identify patterns in the data, incorporating a recurrent connection between hidden layers may reveal significant patterns from past sequences and integrate this information into the current context. Hence, rather than modelling $P\left(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}, ..., \boldsymbol{x}_1\right)$, the RNN approximates:

$$P\left(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}, ..., \boldsymbol{x}_1\right) \approx P\left(\boldsymbol{x}_t | \boldsymbol{H}_{t-1}\right).$$

Figure B.3.1 illustrates the typical RNN architecture in the two ways it appears in academic literature, i.e. as "folded" (LHS) with the typically black square, and unfolded (RHS) network. For instance, consider first the unfolded version. It depicts the feedforward calculation, i.e. from the input sample $\boldsymbol{x}_t$ (bottom) to the prediction $\hat{\boldsymbol{y}}_t$ (top), it follows the same logic and calculations as the MLP. The key difference are the connections between time (left to right). In addition to the FNN, the RNN allows the hidden-to-hidden matrix $\widetilde{\boldsymbol{W}}$ influence the hidden layers at each time step $t = 1, ..., \tau$. The folded presentation of the architecture summarises these inter-time connections by the self-loop.

---

[10]The standard literature assumes a aggregated hidden layer block $\boldsymbol{H}$. For the sake of readability, we assume just one hidden layer, but this could be easily extended.

[11]As a result, it assumes that identical weights can be applied to various time steps, necessitating stationary data. Initially, an hidden state matrix $\boldsymbol{H_0}$ is generated –usually a null-matrix– as well as it corresponding weight matrix $\widetilde{\boldsymbol{W}}$.

Figure B.3.1: Example Computational of a Recurrent Neural Network

Simplistic visualisation of or a plain vanilla folded (LHS) and unfolded (RHS) Recurrent Neural Network with sequential length τ. For the sake of clarity, we choose the box representation of the hidden layer and indicate matrix operations by the filled arrow heads. The black square in the circuit loop on the unfolded graph represents a one-step time delay in the network and summarises the RHS where each node/operation is assigned a time step individually. At the beginning, the model is kicked off by an randomly selected hidden state matrix.
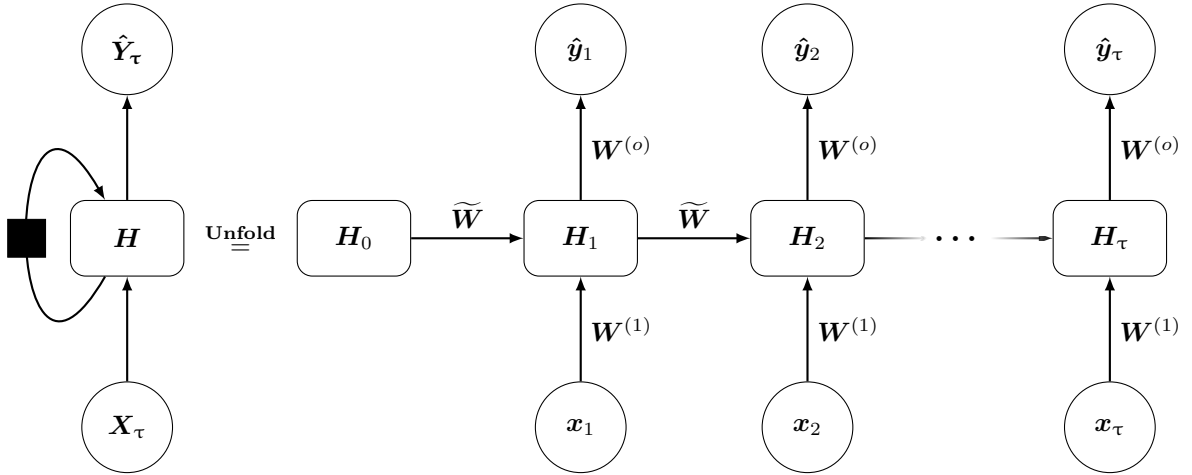
We can express the calculations in Figure B.3.1 adapting the more general matrix notation of Schmidt (2019) by:

$$\boldsymbol{H}_t = \sigma^{(1)}\left(\boldsymbol{X}_t\boldsymbol{W^{(1)}} + \boldsymbol{H}_{t-1}\widetilde{\boldsymbol{W}} + \boldsymbol{b}^{(1)}\right)$$
$$\hat{\boldsymbol{Y}}_t = \sigma^{(o)}\left(\boldsymbol{H}_t\boldsymbol{W}^{(o)}\right).$$

The introduction of time dependency in recurrent neural networks has implications for the learning process. While the gradient descent algorithm remains unchanged, the back-propagation algorithm needs an extension to accommodate the temporal dimension. Two distinct backpropagation algorithms have been developed for this purpose: Real-Time Recurrent Learning (RTRL, Robinson and Fallside (1987)) and Backpropagation Through Time (BPTT, Werbos (1990)). Over time, BPTT has proven to be more efficient in terms of computational resources and simplicity, and thus, we will focus on its application in this context. As seen in Figure B.3.1, the forward pass in an RNN does not differ significantly from that of a standard MLP model. The primary distinction arises during the backward pass. For each input sample $\boldsymbol{x}_t$ there is a prediction $\hat{\boldsymbol{y}}_t$ and associated loss $\ell(\hat{\boldsymbol{y}}_t, \boldsymbol{y}_t)$. We now define a total loss function $\mathcal{L}$ as the sum over all sample losses in a batch:

$$\mathcal{L}\left(\hat{\boldsymbol{Y}}, \boldsymbol{Y}\right) = \sum_{t=1}^{\tau} \ell_t\left(\hat{\boldsymbol{y}}_t, \boldsymbol{y}_t\right). \tag{B.4}$$

The backpropagation through time algorithm considers the total loss and propagates it back through each time step, similar to a simple FNN, and then across the time steps to the beginning. In a way, the errors are travelling back in time to capture the influence of the hidden-to-hidden matrix. Mathematically, this procedure does not differ from (B.2) significantly. In a way, the BPTT algorithm repeatedly and recursively applies the chain rule.

Following the same notation as in Schmidt (2019), we skip some calculations and present the final gradients:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{W}^{(o)}} = \sum_{t=1}^{\tau} \frac{\partial \ell_t}{\partial \hat{\boldsymbol{y}}_t} \cdot \frac{\partial \hat{\boldsymbol{y}}_t}{\partial \sigma^{(o)}} \cdot \boldsymbol{H}_t$$

$$\frac{\partial \mathcal{L}}{\partial \widetilde{\boldsymbol{W}}} = \sum_{t=1}^{\tau} \frac{\partial \ell_t}{\partial \hat{\boldsymbol{y}}_t} \cdot \frac{\partial \hat{\boldsymbol{y}}_t}{\partial \sigma^{(o)}} \cdot \boldsymbol{W}^{(o)} \sum_{k=1}^{t} \frac{\partial \boldsymbol{H}_t}{\partial \boldsymbol{H}_k} \cdot \frac{\partial \boldsymbol{H}_k}{\partial \widetilde{\boldsymbol{W}}}$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{W}^{(1)}} = \sum_{t=1}^{\tau} \frac{\partial \ell_t}{\partial \hat{\boldsymbol{y}}_t} \cdot \frac{\partial \hat{\boldsymbol{y}}_t}{\partial \sigma^{(o)}} \cdot \boldsymbol{W}^{(o)} \sum_{k=1}^{t} \frac{\partial \boldsymbol{H}_t}{\partial \boldsymbol{H}_k} \cdot \frac{\partial \boldsymbol{H}_k}{\partial \boldsymbol{W}^{(1)}}.$$

Schmidt (2019) emphasises that the computation of $\frac{\partial \boldsymbol{H}_t}{\partial \boldsymbol{H}_k}$ is basically a matrix multiplication across the entire sequence τ. The presence of small values (< 1) within this multiplication diminishes the gradient at each time step until it ultimately vanishes, a phenomena widely known as the "vanishing gradient problem". On the other hand, large values ($>> 1$) can result in exceptionally large gradients, also known as the "exploding gradient problem". An optimisation is not possible due to the activation functions' sensitivity to narrow ranges. A rudimentary solution involves gradient clipping, i.e. scaling the gradient similar to the windsorising process. Nevertheless, addressing the vanishing gradient problem requires a more sophisticated approach. Among the most successful implementations is to control actively which information is passed through a layer. These gated RNNs form a distinct subclass of neural networks and promise to solve existing issues. The two most common networks incorporating these information gates are the long short-term memory (LSTM) models (Hochreiter and Schmidhuber, 1997) or the gated recurrent unit (GRU) model (Cho et al., 2014). Notably, LSTMs enjoy considerable popularity among practitioners and are frequently employed for sequential data, which is why our focus is directed towards this model.

Long Short-Term Memory Model

The LSTM is an advanced RNN. As such, its base architecture resembles the standard recurrent neural networks but each neuron is replaced by a memory cell. These memory cells contain an internal state $\boldsymbol{C}_t$ which is a neuron with a self-connected recurrent edge of fixed weight one. This characteristic ensures that the gradient can span numerous time steps without vanishing or exploding. The underlying concept behind this design is relatively straightforward: A plain vanilla RNN maintains its long-term memory in the form of weights which change slowly during training. Additionally, the hold a short-term memory in the form of activation, which pass forward between layers. The LSTM model takes it one step further and introduces an intermediate type of storage via the memory cell (Zhang et al., 2023).

Gates

The internal state of a LSTM memory cell consists of three gates: (i) the "input gate" $\boldsymbol{I}_t$, which decides how much a new input should affect the internal state; (ii) the "forget gate" $\boldsymbol{F}_t$ which decides if and by how much the internal state should be reduced; an (iii) the "output gate" $\boldsymbol{O}_t$ which describes the effect of the neurons output on the internal state. Note the subtle difference: In a MLP, the hidden layer is influenced only by the input and activation functions; in a RNN, the hidden layer is influenced by the input, activation function and the (constant) hidden layer of the previous time-step; the LSTM allows the previous hidden layer

to be kept, gated, or ignored before including it in the cell state. The mathematics behind these gates are basic matrix multiplications. Keeping the same notation as in Section B.3, we denote an output gate $\boldsymbol{O}_t$, an input gate $\boldsymbol{I}_t$ and a forget gate $\boldsymbol{F}_t$ and define them as:

$$\boldsymbol{O}_t = \text{sig}\left(\boldsymbol{X}_t\boldsymbol{W}^{(xo)} + \boldsymbol{H}_{t-1}\boldsymbol{W}^{(ho)} + \boldsymbol{b}^{(o)}\right)$$

$$\boldsymbol{I}_t = \text{sig}\left(\boldsymbol{X}_t\boldsymbol{W}^{(xi)} + \boldsymbol{H}_{t-1}\boldsymbol{W}^{(hi)} + \boldsymbol{b}^{(i)}\right)$$

$$\boldsymbol{F}_t = \text{sig}\left(\boldsymbol{X}_t\boldsymbol{W}^{(xf)} + \boldsymbol{H}_{t-1}\boldsymbol{W}^{(hf)} + \boldsymbol{b}^{(f)}\right),$$

where $\text{sig}(\cdot)$ denotes the sigmoid function, $\boldsymbol{W}^{(xo)}$, $\boldsymbol{W}^{(xi)}$, $\boldsymbol{W}^{(xf)} \in \mathbb{R}^{(d \times N_1)}$ and $\boldsymbol{W}^{(ho)}$, $\boldsymbol{W}^{(hi)}$, $\boldsymbol{W}^{(hf)} \in \mathbb{R}^{(N_1 \times N_1)}$ are weight matrices with their respective biases $\boldsymbol{b} \in \mathbb{R}^{(1 \times N_1)}$.

Next, we define an "input node" $\widetilde{\boldsymbol{C}}_t \in \mathbb{R}^{(n \times N_1)}$. This node is a pre-filter and will later update the internal state of the cell. It considers the input to the neuron and passes it through a hyperbolic tan function which limits the output to $(-1, 1)$. The input node is defined as:

$$\widetilde{\boldsymbol{C}}_t = \tanh\left(\boldsymbol{X}_t\boldsymbol{W}^{(xc)} + \boldsymbol{H}_{t-1}\boldsymbol{W}^{(hc)} + \boldsymbol{b}^{(c)}\right),$$

with its respective weights $\boldsymbol{W}^{(xc)} \in \mathbb{R}^{(d \times N_1)}$, $\boldsymbol{W}^{(hc)} \in \mathbb{R}^{(N_1 \times N_1)}$ and bias $\boldsymbol{b}^{(c)} \in \mathbb{R}^{(1 \times N_1)}$.

Figure B.3.2 illustrates the first step of the LSTM processing procedure. Similar to an RNN, the input data, along with the hidden state of the previous time step feeds into the cell. Then, the combined information is passed through four distinct gates. These employ the sigmoid (forget, input and output gate) activation function, limiting the output within the range $(0, 1$, while the input node passes through a hyperbolic tangent activation function, clipping the output on an $(-1, 1)$ interval. Note that the calculation throughout the three gates differs only on their weights and bias.

Figure B.3.2: Visualisation of a LSTM Cell (I)

Visual representation of a LSTM model. This figure is copied from Schmidt (2019) and Zhang et al. (2023). All credits belong to the authors. For a detailed description of changes between steps, please see Zhang et al. (2023) or the Book Website.
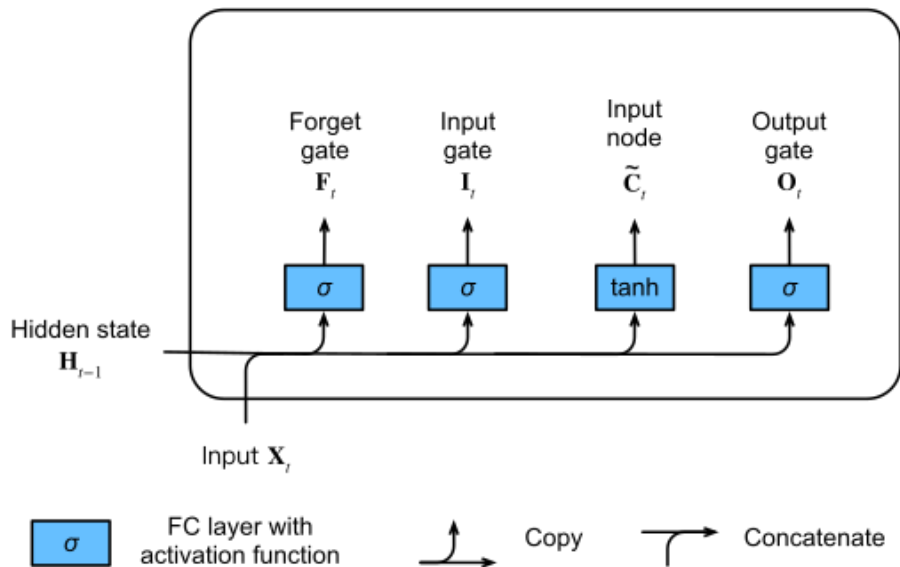
Internal State & Output

In the second step, e gate outputs determine and influence the current cell state and neuron output. We introduce the internal cell state, denoted as $\boldsymbol{C}_t \in \mathbb{R}^{n \times N_1}$ which regulates the preservation of prior information before passing it to the next cell state. Here, the forget gate determines the extent to which the previous state should be discarded, while the input gate and node dictate the degree to which new input should affect the cell state. Note that this provides flexibility, enabling the learning mechanism to decide when to maintain the value unaltered and when to modify it, based upon the incoming data. The mathematical computation at this stage is as follows:

$$\boldsymbol{C}_t = \boldsymbol{F}_t \odot \boldsymbol{C}_{t-1} + \boldsymbol{I}_t \odot \widetilde{\boldsymbol{C}_t},$$

where $\odot$ denotes the element-wise multiplication or Hadamard product. At this point, the update of the internal state is complete and will be transmitted to the next time step. To calculate the output of the LSTM cell, i.e. the hidden state $\boldsymbol{H}_t$, the model employs the tanh function on the internal memory state *after* the forget and input updates and subsequently multiplies the output point-wise with the results of the output gate:

$$\boldsymbol{H}_t = \boldsymbol{O}_t \odot \tanh\left(\boldsymbol{C}_t\right).$$

Figure B.3.3 summarises the internal process of a LSTM cell and combines the steps discussed. The mathematical operations are presented in light blue, while the activation functions are given in dark blue rectangles. Input data, along with the previous hidden state is passed through gates which then update the internal cell state. This internal cell state is passed on the next time step and influences the output of the cell. This procedure is repeated at the next time step for all samples in the batch.

Figure B.3.3: Visualisation of a LSTM Cell (II)

Visual representation of a LSTM model. This figure is copied from Schmidt (2019) and Zhang et al. (2023). All credits belong to the authors. For a detailed description of changes between steps, please see Zhang et al. (2023) or the Book Website.
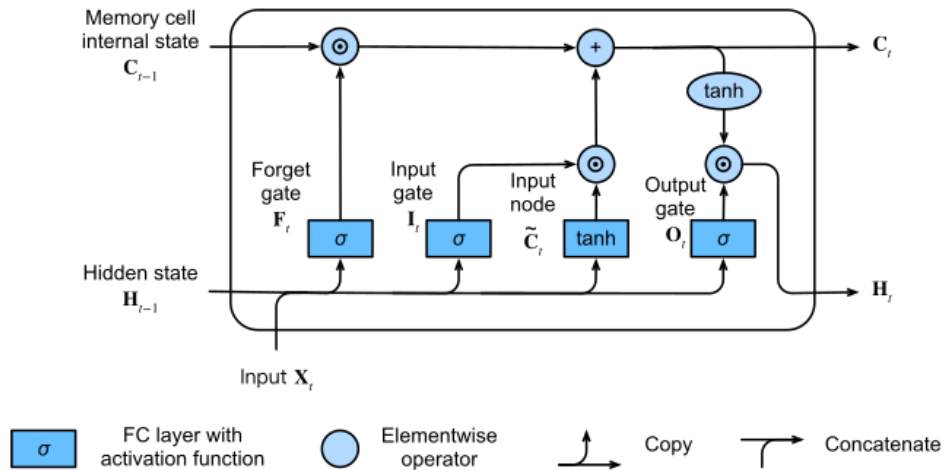
APPENDIX C

IMPLIED VOLATILITY DYNAMICS

C.1 LSTM Model Specification

We discussed the concept of data pre-processing earlier in Chapter B. After ensuring the stationarity of each beta time series, we compare different scaling methods on our dataset, i.e. standardise each time series on [0, 1] or [-1, 1] using the `preprocessing` library, normalise the data or leave it unchanged. Out of these four data manipulations, the standardisation on [0, 1] yields the best result on the validation set. Less surprising, the raw data performance comes in last. Unfortunately, determine the optimal hyperparameters for the model f remains a challenging task without a general go-to solution. The optimal choice of parameters, i.e. number of lags p, the batch size B, number of nodes n, epochs E and hidden layers d as well as the activation function requires a trail-and-error approach and is often considered more an art then science. For this, we set:

$$p \in [3, 9, 21, 42, 90, 102, 135], \qquad B \in [10, 32, 64, 128, 250],$$
$$E \in [50, 100, 150, 200, 250, 300], \qquad n \in [10, 50, 80, 100, 150, 200],$$

and examine all possible permutations of the hyperparameter through a grid search.[1] Overall, we compared 1,250 different models and measure the best model by its MSE and the R^2 value on the validation set.

From the outset, it was evident that the primary issue with the data was its insufficient size, leading to overfitting of the model and resulting exceptionally low training loss, but divergent validation loss. To overcome this data limitation, we follow Srivastava et al. (2014) and install the `DropOut` technique as a regulation mechanism into each layer of our model, which randomly disables 5% of the connections. Furthermore, we explore the option of using the `DropOut` technique as a Bayesian approximation as discussed in Chapter B. In fact, our

---

[1]A grid search describes an algorithm in which any possible permutation is tested, validated and the results archived. The final result describes the best performing model hyperparameter based on the chosen loss function. For this we utilised two machines: (i) a MacBook Pro with 2GHz Quad-Core Intel Core i5 and 16GB 3733 MHz LPDDR4X RAM; and (ii) a Dell XPS 13 with 1.3GHz 8×Intel Core i7 and 16GB LPDDR4X RAM. We used the `multiprocessing` function of Python. This enables cores to run parallel and evaluate model parallel. For our grid-search, we used twelve physical and four virtual cores over a period of two weeks to obtain optimal results. We test different permutations with one or two hidden layer and find best results for two hidden layers.

investigation reveals that the use of the Monte Carlo Dropout function enhances the model's performance, as measured by MSE and R^2.[2] Figure C.1.1 illustrates the use of the Dropout as Bayesian approximation, i.e. we generate 1000 predictive simulations of first parameter β_0 for $t = 01/08/2022$ at 16:00 GMT and plot the kernel density of the forecasts. We already inverted the scaling, i.e. the x-axis represents the true β_0 values. The results cluster around -0.33 which indicates a general implied volatility level of 71.5%, during a time when the 10-day ATM volatility was 70.5%. Furthermore, there is some wider variation and a small "bump" at -0.615 indicating a volatility level of 54%. This is a consequence of the relatively stable underlying bitcoin price at \$22,500 for the two months leading up to August and a generally declining implied volatility. This bump may indicate the future trend and direction of the volatility. This type of probability density function paves the way for a new strain of option pricing research that we will not further pursue in this work.

Figure C.1.1: Kernel Density Estimate

Results of 1000 simulations of the β_0 coefficient on 1 August 2022 at 08:00 GMT for 1 August 2022 at 16:00 GMT, presented as a kernel density estimate. The x-axis represents the predicted value, while the y-axis represents the estimated probability density.


---
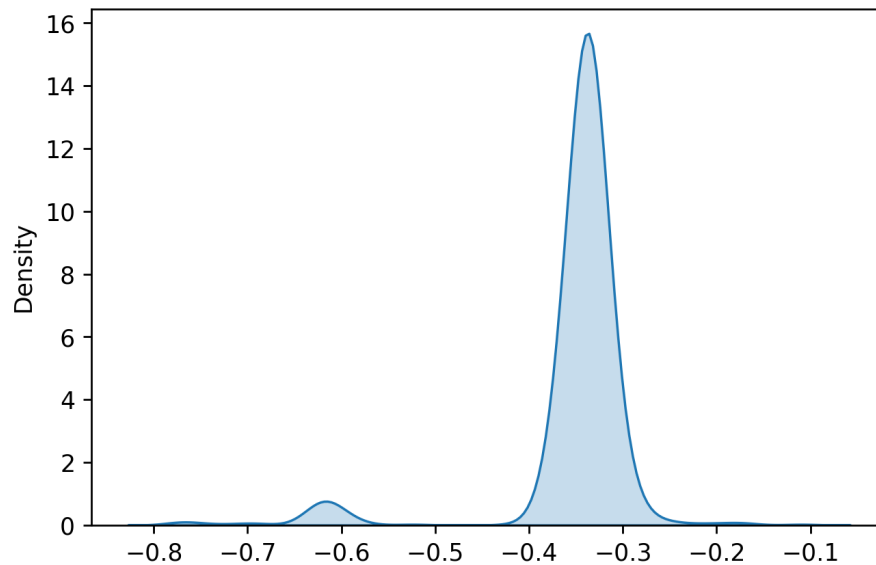
[2]Using the same hyperparemeter, the the Monte Carolo approach yiels and average improvement of 6% (4%) in terms of MSE (R^2) on the validation set.

Figure C.1.2 depicts the performance of our LSTM model across epochs in terms of the MSE loss function. The plot comprises two curves: the training loss and the validation loss. The training and validation loss curves both exhibit a steady decline over epochs, indicating that the model is learning effectively from the training data and generalising well to the validation data. The convergence of these curves implies that the model achieves excellent results with minimal over- or underfitting.[3]

Figure C.1.2: Comparison of Training Loss and Validation Loss of LSTM Model

Comparison of training loss (black) and validation loss (gold) of the LSTM model with $n = 150$ nodes, $p = 102$ lags, batch size $B = 64$ and two hidden layer. The x-axis represents the epoch number, while the y-axis represents the MSE value.

---
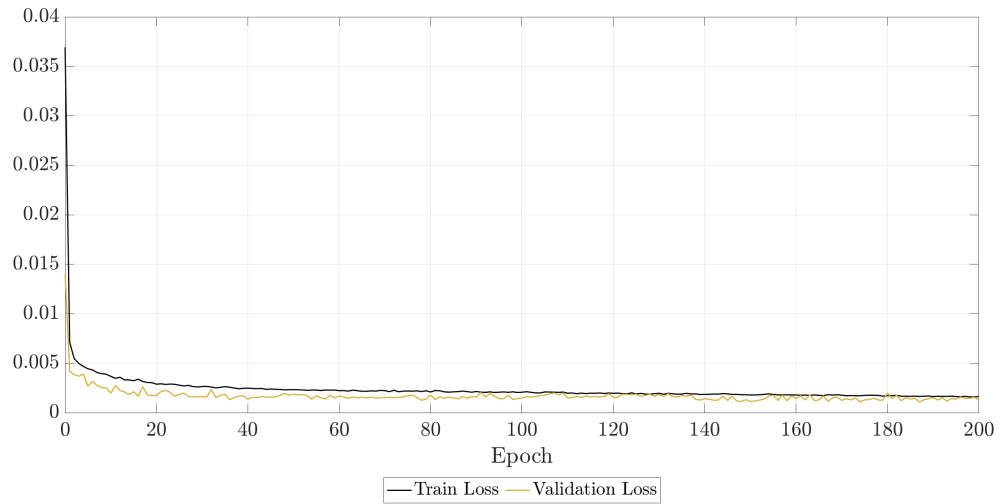
[3]Note the atypical observation, as the validation loss is lower than the training loss. This may be attributed to two factors. First, the `DropOut` function is applied only during the training phase, and not during prediction. During training, a random 5% of connections are disabled, while all connections are active during prediction. Secondly, it could be because the training loss is measured during the training phase, while the validation loss is measured after each epoch, thereby leading to discrepancies in their magnitudes.

C.2   Robustness Check

Figure C.2.1: Model 2 Out-of-Sample Time Series β Coefficients

Time series of each β coefficient (red) of the deterministic implied volatility surface model estimated using the robust ordinary least square method on Model 2, covering our out-of-sample period from August 2022 to December 2023 in eight hour steps as well as the step-ahead prediction using the random walk benchmark (purple), vector autoregressive model (orange) and Long-Short-Term memory model (green). We omit the horizontal time axis for the sake of clearly but keep it for the last panel. The vertical axis is unit-free and changes for each parameter.
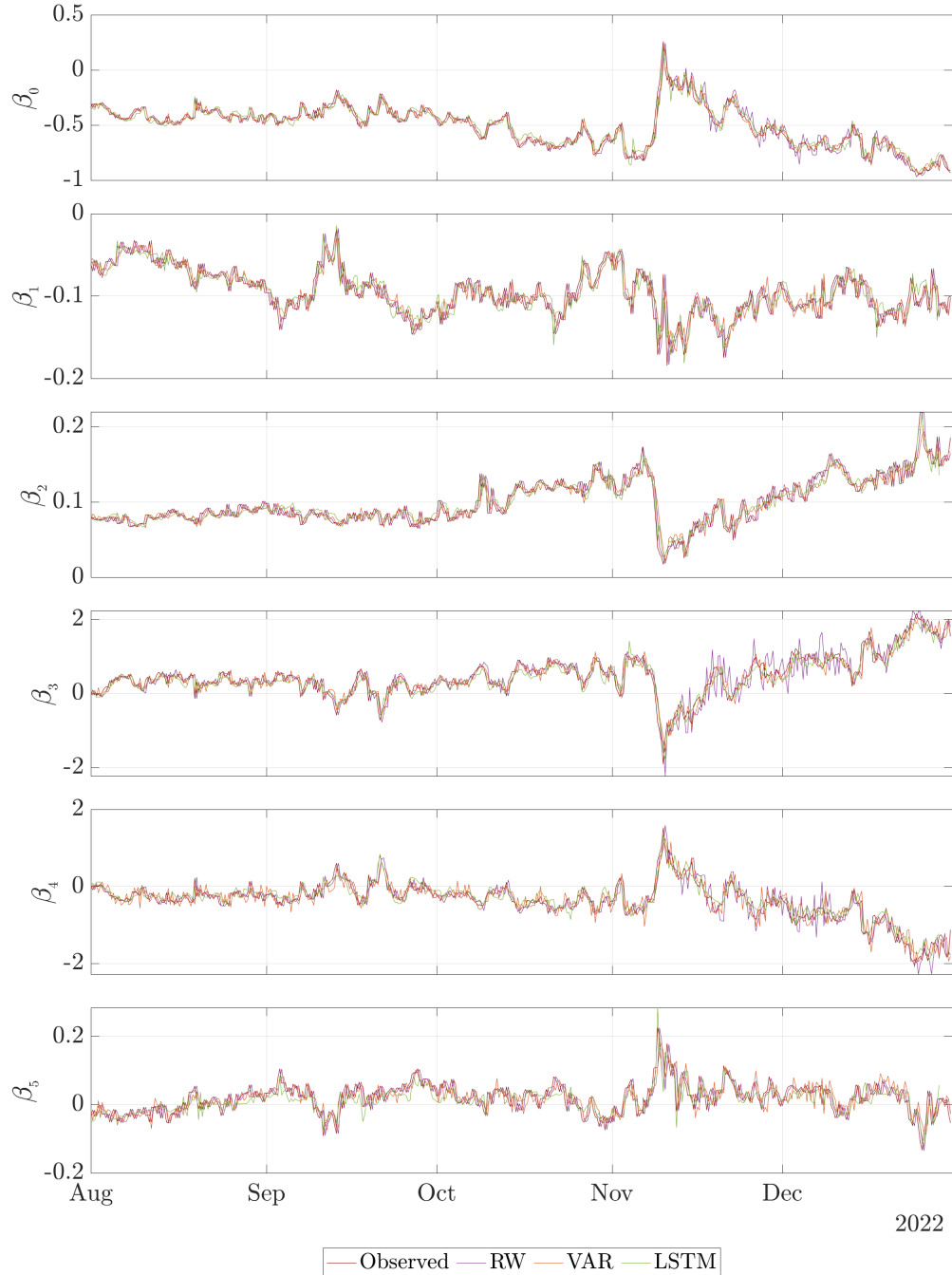
Table C.2.1: Model 2 Implied Volatility Out-of-Sample Average Prediction Errors by Moneyness and Maturity

Comparison of out-of-sample prediction errors using a variety of models, expressed as average of the three performance measures, i.e. Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Mean Directional Error (MDE). We consider the implied volatility surface backed out of the predicted beta coefficients for the period from August 2022 to December 2022 on an eight hour frequency and divide the results for different maturities (Short Term: 3-14 days; Mid Term: 14-30 days; Long-Term: 30+ days), moneyness levels (OTM Put: moneyness between 0.5-0.94; ATM: moneyness between 0.94-1.06; OTM Call: moneyness between 1.06-2) as well as any permutation of these two. The last row and column depict the overall prediction errors. We highlight the overall best results in blue.

	Short-Term			Mid-Term			Long-Term			Total		
	RMSE	MAE	MCD	RMSE	MAE	MCD	RMSE	MAE	MCD	RMSE	MAE	MCD
OTM Put:												
RW	0.0735	0.0436	50.02%	0.0645	0.0417	51.42%	0.0763	0.0449	52.48%	0.0738	0.0441	52.21%
VAR	0.0756	0.0432	54.56%	0.0603	0.0398	52.84%	0.0391	0.0283	53.43%	0.0502	0.0326	53.47%
LSTM	0.0746	0.0442	58.53%	0.0649	0.0422	57%	0.0514	0.0392	53.24%	0.0579	0.0405	54.7%
ATM:												
RW	0.0463	0.0349	50.34%	0.0388	0.0278	45.05%	0.0653	0.0338	59.94%	0.0542	0.0327	48.85%
VAR	0.0443	0.034	53.57%	0.0338	0.0247	46.79%	0.0231	0.0166	59.94%	0.0334	0.0239	50.26%
LSTM	0.0441	0.0344	58.49%	0.0344	0.0266	53.59%	0.0368	0.0263	51.53%	0.0387	0.0289	54.15%
OTM Call:												
RW	0.0524	0.0379	49.2%	0.0458	0.0322	54.13%	0.074	0.0401	50.05%	0.0698	0.039	49.77%
VAR	0.048	0.0355	49.81%	0.0401	0.029	46.63%	0.0287	0.0212	49.84%	0.0321	0.0232	49.46%
LSTM	0.0431	0.0335	56.85%	0.0376	0.0281	55.43%	0.0401	0.0294	52.39%	0.04	0.0296	53.09%
Total:												
RW	0.0574	0.0383	50.59%	0.0517	0.0343	47.44%	0.0753	0.0418	51.08%	0.0698	0.0401	50.44%
VAR	0.0567	0.0372	52.96%	0.047	0.0317	48.98%	0.0321	0.0231	50.94%	0.0389	0.0264	50.91%
LSTM	0.0553	0.0372	58.%	0.0488	0.0329	55.43%	0.0446	0.033	52.61%	0.0468	0.0336	53.81%

Table C.2.2: Model 2 Diebold-Mariano Test Statistic

Pair-wise comparison of the Diebold and Mariano (1995) (DM) statistic for out-of-sample predictions from August 2022 to December 2022. We divide the results for different maturities (Short Term: 3-14 days; Mid Term: 14-30 days; Long-Term: 30+ days), moneyness levels (OTM Put: moneyness between 0.5-0.94; ATM: moneyness between 0.94-1.06; OTM Call: moneyness between 1.06-2) as well as any permutation of these two. The DM test compares the forecasting accuracy of the RW/VAR (column) with the VAR/LSTM (row) model. It tests the null hypothesis that the forecasts have the same accuracy, i.e. a positive DM statistic indicates that the first model has larger average forecast error compared to the second and hence the second model is preferred. The p-value of the statistic is expressed using one (three) asterisks (***) for the 10% (1%) significance level.

	Short-Term		Mid-Term		Long-Term		Total	
	VAR	**LSTM**	**VAR**	**LSTM**	**VAR**	**LSTM**	**VAR**	**LSTM**
OTM Put:								
RW	-0.65	-0.42	2.81^{**}	-0.2	10.31^{***}	7.6^{***}	10.12^{***}	7.19^{***}
VAR		0.32		-2.22		-30.44^{***}		-9.42^{***}
ATM:								
RW	4.49^{***}	3.04^{***}	9.86^{***}	5.09^{***}	7.13^{***}	5.55^{***}	7.71^{***}	6.07^{***}
VAR		-0.2		-0.84		-18.4^{***}		-14.12^{***}
OTM Call:								
RW	5.51^{***}	6.55^{***}	9.79^{***}	6.5^{***}	15.8^{***}	12.94^{***}	15.93^{***}	13.52^{***}
VAR		4.52^{***}		2.45^{**}		-32.82^{***}		-25.07^{***}
Total:								
RW	0.56	1.89^{*}	6.41^{***}	2.52^{**}	20.96^{***}	16.64^{***}	2.123^{***}	16.92^{***}
VAR		0.99		-1.66^{**}		-53.01^{***}		-22.6^{***}

Table C.2.3: Model 2 Trading Strategy Evaluation

Summary statistics of the three trading strategies (straddle, delta hedge and buy-and-hold) for short-term ATM options based on trading signals generated by distinct predictive models (RW, VAR and LSTM). The performance measures include the average return, standard deviation, resulting Sharpe Ratio, and Sortino Ratio, all expressed in percentage terms for the out-of-sample period August to December 2022 using the intraday eight hour rebalancing frequency. We have highlighted the best model in blue.

	Mean	Std. Dev.	Median	Sharpe Ratio	Sortino Ratio
Strategy A *(ATM Straddle)*					
RW	-0.36	6.16	-0.99	-5.91	-8.92
VAR	-0.1	6.05	-0.88	-1.57	-2.55
LSTM	0.49	6.06	-0.08	8.16	13.93
Strategy B *(delta-hedged portfolio)*					
RW	-3.02	21.67	-0.31	-13.92	-15.91
VAR	-2.15	18.23	-0.24	-11.79	-13.75
LSTM	0.23	22.55	0	1.02	1.4
Strategy C *(Buy and Hodl)*	-0.94	28.1	-1.07	-4.54	-5.6