ISSN: 3079-9325

Transformer-Based Architecture for Real-Time Option Greeks Estimation Under Extreme Market Conditions

Liang Zhao[1,*], Mei Lin[1], and Hiroshi Tanaka[2]

[1] Department of Computer and Information Science and Engineering, University of Florida, USA

[2] Graduate School of Economics, Kyoto University, Japan

[*] Corresponding Author: liang.zh877@outlook.com

Abstract

The accurate and timely estimation of option Greeks remains a critical challenge in financial risk management, particularly during periods of extreme market volatility when traditional computational methods encounter severe limitations in both speed and reliability. This paper presents a novel application of Transformer-based deep learning architectures to the problem of real-time option Greeks estimation under extreme market conditions, addressing fundamental challenges that have constrained conventional approaches including computational bottlenecks, numerical instability, and inadequate handling of long-range temporal dependencies in volatility dynamics. We develop a specialized attention mechanism that exploits the structural properties of option surfaces while maintaining computational efficiency through strategic architectural design incorporating multi-head self-attention, gated neural network mechanisms that enforce economic rationality constraints, and positional encoding adapted for financial time series exhibiting non-stationary behavior. The empirical investigation employs comprehensive datasets spanning multiple market regimes including the 2008 financial crisis characterized by VIX levels exceeding 80 percent as documented in detailed intraday records, the August 2015 volatility spike reaching 53 percent, and the March 2020 COVID-19 pandemic market disruption with VIX peaking at 89.53 percent, providing robust assessment across diverse stress scenarios that reveal the catastrophic failure modes of traditional methods. Our Transformer-based approach achieves Delta estimation accuracy with Mean Absolute Error below 0.001 for at-the-money options during normal market conditions and maintains stable performance with MAE below 0.002 during extreme volatility events where traditional finite difference methods exhibit errors exceeding 0.05, representing more than twentyfivefold improvement in accuracy under stress conditions. The architecture leverages a stratified training strategy that oversamples extreme volatility regimes by factors exceeding thirteen times their natural occurrence frequency, ensuring robust generalization to crisis scenarios despite their rarity in historical data comprising less than two percent of trading days. Furthermore, the architecture delivers inference latency below 100 microseconds per option contract on modern GPU hardware, enabling genuine real-time Greeks calculation for large portfolios containing thousands of positions that require continuous hedging adjustments as volatility surfaces shift rapidly during market stress. This research establishes Transformer models as a transformative methodology for derivatives risk management, offering practitioners a robust tool for maintaining accurate hedge ratios and risk metrics even during the most turbulent market periods when precise Greeks estimation proves most critical for portfolio survival.

ISSN: 3079-9325

Keywords

Transformer Architecture, Attention Mechanism, Option Greeks, Delta Hedging, Volatility Surface, Extreme Market Conditions, Deep Learning, Real-Time Computation, Financial Risk Management, VIX Index, Market Crisis, Gated Neural Networks, Stratified Sampling

1. Introduction

The computation of option Greeks, representing the sensitivities of derivative prices to various underlying factors including asset price movements, volatility fluctuations, time decay, and interest rate changes, constitutes one of the most fundamental yet computationally demanding tasks in quantitative finance[1]. These sensitivities directly underpin portfolio hedging strategies that protect multi-billion dollar derivative books from adverse market movements, risk limit monitoring systems that prevent catastrophic losses by flagging excessive exposures before they materialize, and regulatory capital calculations that determine the financial resources institutions must hold against potential losses[2]. Traditional methodologies for Greeks estimation, including analytical differentiation of closed-form pricing formulas where available and numerical finite difference approximations for complex instruments lacking analytical solutions, have served the financial industry for decades as the backbone of risk management infrastructure[3]. However, these conventional approaches encounter increasing strain under modern market conditions characterized by unprecedented portfolio complexity with thousands of interdependent positions, algorithmic trading speed requirements demanding microsecond-level responsiveness, and episodic volatility spikes that stress computational frameworks beyond their design limits[4].

The challenge intensifies dramatically during periods of extreme market stress, creating a cruel paradox where accurate Greeks become most critical for survival precisely when the methods relied upon for their computation exhibit catastrophic degradation in both accuracy and reliability. The 2008 financial crisis provided a stark demonstration of these limitations, as the Chicago Board Options Exchange Volatility Index (VIX), commonly known as the fear gauge measuring market expectations of 30-day volatility implied by S&P 500 index options, surged from typical levels around 20 percent during calm periods to a record closing value of 80.74 percent on November 21, 2008[5]. This quadrupling of expected market volatility within months created unprecedented challenges for risk management systems as option portfolios experienced violent daily swings in value driven by rapidly changing Greeks, while the very numerical methods relied upon for risk calculation became unreliable due to multiple compounding factors including widening bid-ask spreads that corrupted input data quality, breakdown of continuous hedging assumptions as markets gapped discontinuously with trading halts and circuit breakers triggering, and computational resource exhaustion as systems struggled to revalue thousands of positions fast enough to maintain meaningful real-time risk metrics[6].

The pattern repeats with each subsequent market dislocation, confirming that extreme volatility represents not a rare aberration but a recurring feature of financial markets that risk management systems must reliably handle. The August 2015 volatility spike saw the VIX briefly touch 53.29 percent following Chinese equity market turmoil and concerns about economic slowdown, creating another period where traditional Greeks computation methods struggled to maintain accuracy as implied volatility surfaces exhibited dramatic shifts in skew and term structure that violated the smooth variation assumptions underlying finite

ISSN: 3079-9325

difference approximations[7]. The March 2020 COVID-19 pandemic onset witnessed even more extreme conditions as global uncertainty about the virus's economic impact drove the VIX to an all-time intraday high of 89.53 percent on March 16, 2020, exceeding even the 2008 crisis peak and creating the most challenging Greeks estimation environment in modern financial history where bid-ask spreads widened to levels making market data nearly unusable for precise calculations[8].

Traditional finite difference methods for Greeks computation, the industry workhorse approach for instruments lacking closed-form solutions, evaluate option pricing functions at perturbed input values and approximate derivatives through ratios of price changes to input perturbations[9]. For Delta, measuring sensitivity to underlying asset price changes, a standard two-point centered difference formula evaluates prices at the current spot price plus and minus a small increment, differences these prices, and divides by twice the increment to obtain an approximation to the first derivative. While conceptually straightforward and applicable to arbitrary pricing models implementable as computable functions, this approach suffers from multiple fundamental weaknesses that become acute under stress conditions[10]. The method requires multiple expensive pricing evaluations per Greek, with second-order sensitivities like Gamma requiring four or more pricing function calls creating severe computational burdens for portfolios containing thousands of options requiring simultaneous Greeks calculation multiple times daily[11]. The perturbation size selection presents an intractable tradeoff, with large perturbations introducing truncation error as the finite difference deviates from the true derivative due to the nonlinear curvature of option value functions, while small perturbations amplify catastrophic cancellation errors as the price difference becomes comparable to floating point precision limits, a dilemma lacking satisfactory universal resolution particularly for options exhibiting discontinuous behavior near barriers or kinks in payoff functions[12].

The emergence of deep learning as a transformative force across pattern recognition domains, achieving superhuman performance in tasks ranging from image classification and object detection to natural language understanding and machine translation, naturally suggests its application to financial computation problems that exhibit complex nonlinear patterns amenable to data-driven learning rather than requiring explicit algorithmic specification[13]. Early applications of neural networks to option pricing date to the 1990s, with pioneering work by Hutchinson, Lo, and Poggio demonstrating feasibility of learning pricing functions from simulated data, but these efforts primarily targeted pricing rather than Greeks estimation and employed relatively simple feedforward architectures lacking sophisticated mechanisms for capturing the rich structural relationships characterizing option surfaces across strikes and maturities[14]. The introduction of the Transformer architecture by Vaswani and colleagues in 2017, originally motivated by machine translation tasks requiring attention to long-range dependencies between words in sentences separated by many tokens, marked a fundamental paradigm shift enabling models to dynamically weight the relevance of different input elements through learned attention mechanisms rather than processing information sequentially as recurrent networks do or with fixed local receptive fields as convolutional architectures employ[15].

This paper investigates the application of Transformer-based architectures specifically designed for the problem of real-time option Greeks estimation under extreme market conditions, addressing simultaneously the computational efficiency requirements of production trading systems demanding sub-millisecond latency and the accuracy challenges posed by volatile markets where traditional numerical methods experience catastrophic

ISSN: 3079-9325

failure precisely when reliability matters most. We develop specialized network designs that incorporate financial domain knowledge through architectural constraints encoding no-arbitrage principles, adapted attention mechanisms that efficiently process option surface structure by identifying relevant cross-strike and cross-maturity relationships, and training strategies that ensure robust generalization across diverse market regimes including rare stress events dramatically underrepresented in historical data. The investigation employs comprehensive datasets spanning normal market conditions for baseline assessment establishing competitive performance against traditional methods under benign conditions, alongside detailed crisis period data including minute-by-minute records from October-November 2008, August 2015, and March 2020 volatility spikes enabling rigorous stress testing of model behavior precisely when accuracy matters most for preventing catastrophic portfolio losses.

The motivation for this research stems from pressing practical needs facing risk management infrastructure at financial institutions where derivative portfolio values measured in billions of dollars demand continuous monitoring through Greeks that must be calculated thousands of times daily as markets move and volatility surfaces shift, with even minor inaccuracies in Greeks potentially translating to tens of millions in hedging errors or undetected risk exposures. While traditional methods remain viable under benign market conditions with stable volatility and liquid markets providing reliable pricing data, their systematic failure modes during stress create unacceptable operational risks as precisely the moments requiring most careful risk management and accurate hedging coincide with computational framework breakdown. Deep learning approaches offering robust performance across all market conditions from calm to crisis while delivering evaluation speeds enabling genuine real-time Greeks calculation represent transformative advances with clear practical value for institutions whose survival during the next crisis may depend on maintaining accurate risk metrics when others cannot. From theoretical perspectives, exploring how Transformer attention mechanisms naturally capture option surface structure, whether financial time series exhibit long-range dependencies amenable to this architecture, and how economic constraints can be embedded into neural designs provides insights relevant to broader questions about appropriate machine learning methodologies for financial applications where reliability, interpretability, and handling of rare extreme events prove as important as average-case accuracy.

2. Literature Review

The literature on option Greeks computation has evolved over several decades from early analytical formulas applicable to simple instruments through increasingly sophisticated numerical methods addressing complex derivatives, with recent years witnessing accelerating interest in machine learning approaches offering potential advantages in both computational speed and accuracy under challenging conditions[16]. The foundational Black-Scholes-Merton framework published in 1973 provided not only closed-form option pricing formulas revolutionizing derivatives markets by enabling systematic valuation, but also analytical expressions for Greeks including Delta, Gamma, Vega, Theta, and Rho computed through straightforward differentiation of the pricing formula with respect to relevant parameters[17]. For vanilla European options on non-dividend-paying stocks satisfying the model's assumptions of constant volatility and continuous trading, these formulas enable exact Greeks computation requiring only evaluation of the cumulative standard normal distribution function and elementary algebraic operations, offering computational efficiency measured in microseconds and numerical precision limited only by floating-point arithmetic that

ISSN: 3079-9325

established a gold standard against which alternative methods must be measured for both speed and accuracy[18].

However, the applicability of analytical Greeks remains severely constrained, limited to the narrow class of instrument types and market assumptions for which closed-form formulas have been derived through mathematical analysis[19]. American options exercisable at any time prior to maturity lack closed-form solutions except in special cases, exotic derivatives with path-dependent payoffs or barrier features generally require numerical methods, and realistic models incorporating stochastic volatility to capture observed volatility smiles or jump processes to model discontinuous price movements typically necessitate Monte Carlo simulation or partial differential equation solution techniques for which Greeks computation presents additional challenges[20]. Finite difference methods emerged as the workhorse computational approach, numerically approximating derivatives by evaluating pricing functions at perturbed input values and computing difference quotients that converge to true derivatives as perturbation sizes shrink. While conceptually simple, universally applicable to any pricing model implementable as a computable function, and straightforward to implement requiring only multiple calls to existing pricing routines, finite difference methods face well-documented challenges that become acute in practice including computational cost scaling linearly with the number of Greeks required as each sensitivity demands separate perturbations, numerical instability particularly for higher-order derivatives where errors compound through multiple differencing operations, and fundamental difficulty selecting appropriate perturbation sizes that balance competing concerns of truncation error from finite difference approximations versus roundoff error amplification as price differences approach machine precision[21].

Alternative Greeks computation approaches have been developed addressing some finite difference limitations while introducing their own constraints and applicability conditions. The pathwise differentiation method, also known as the likelihood ratio or score function method in statistical contexts, offers an approach for Greeks computation within Monte Carlo simulation frameworks by differentiating simulated payoffs directly with respect to parameters of interest rather than perturbing inputs and differencing prices[22]. When applicable, pathwise methods often exhibit substantially lower variance than finite difference approaches for the same computational budget, providing more accurate Greeks estimates from a given number of simulation paths[23]. However, applicability requires payoff differentiability with respect to the sensitivity parameter, excluding certain exotic options with discontinuous payoffs at barriers or exercise boundaries where the derivative fails to exist in classical sense[24]. Adjoint algorithmic differentiation provides another sophisticated approach, automatically generating efficient code for gradient computation by systematically applying the chain rule throughout the computational graph defining the pricing function, offering potential for computing all parameter sensitivities simultaneously with computational cost comparable to a single function evaluation rather than scaling linearly with parameter count. This dramatic efficiency advantage for high-dimensional parameter spaces has driven adoption in some quantitative finance applications, though implementation complexity and software tooling requirements have limited widespread deployment compared to the simplicity of finite difference methods requiring only existing pricing code[25].

The application of neural networks to option pricing problems began gaining traction in the 1990s as computational capabilities advanced sufficiently to train networks on realistic financial datasets rather than toy problems[26]. Hutchinson, Lo, and Poggio's influential 1994

ISSN: 3079-9325

study demonstrated that multilayer feedforward networks could learn to approximate Black-Scholes prices from simulated training data without explicit knowledge of the closed-form pricing formula, establishing the feasibility principle that neural networks could extract complex pricing relationships from data through pattern recognition rather than requiring human derivation of mathematical formulas[27]. This work extended to Greeks estimation by training separate networks on computed derivative values, showing that learned Greeks approximations could match or exceed the accuracy of finite difference methods on test data[28]. While these results primarily demonstrated capability rather than clear practical advantages over analytical methods for the Black-Scholes case where exact formulas exist, they established important principles including that neural networks could discover pricing patterns from data without explicit model specification and that learned approximations could potentially avoid the numerical instabilities plaguing finite difference methods[29].

Subsequent research through the 2000s and early 2010s explored various neural network architectures and training strategies for option pricing and Greeks computation, with mixed results that generated both enthusiasm about machine learning's potential and skepticism about whether the complexity and opaqueness of neural approaches justified adoption when traditional methods worked adequately under normal conditions[30]. Some studies reported superior out-of-sample pricing accuracy for networks compared to misspecified parametric models, though careful interpretation requires distinguishing networks' function approximation capability from their appropriateness as fundamental pricing models versus empirical curve-fitting tools. The reliability concerns proved particularly acute for Greeks estimation where small errors in learned pricing functions can amplify into large derivative approximation errors, and where the lack of transparency in neural network computations complicated validation and debugging compared to traditional methods with clear mathematical foundations[31].

More recent work has specifically targeted Greeks estimation through neural approaches, motivated by both potential speed advantages from fast neural network inference after expensive offline training and the possibility that networks might learn stable derivative approximations avoiding the numerical issues plaguing finite difference methods[32]. Research has explored two main approaches with distinct advantages and challenges. Direct methods train networks to predict Greeks directly from option characteristics and market conditions, treating Greeks estimation as a supervised regression problem where training targets are Greeks values computed through alternative benchmark methods. This approach potentially avoids compounding approximation errors from pricing function learning with additional errors from numerical differentiation, instead learning the derivative function directly as the target pattern to recognize[33]. However, it requires availability of accurate training labels which themselves must be computed somehow, typically through expensive but reliable methods like Monte Carlo with variance reduction, creating a bootstrap problem where training the fast network requires extensive application of slow traditional methods. Alternative indirect approaches train networks on pricing functions then extract Greeks through differentiation of the trained network, leveraging the smooth and continuously differentiable nature of neural network architectures to potentially provide more stable derivatives than the underlying pricing function exhibits, though this compounds learning errors from pricing approximation with differentiation approximation[34].

A particularly relevant development emerged from Yang, Zheng, and Hospedales' 2017 paper introducing gated neural networks for option pricing that incorporate economic constraints including no-arbitrage principles directly into network architecture through carefully

ISSN: 3079-9325

designed gating mechanisms and activation functions[35]. Their approach, termed rational by design, ensures that learned pricing functions automatically satisfy fundamental economic properties that option prices must obey regardless of market conditions, including monotonicity constraints requiring call option values to decrease with strike price, appropriate asymptotic behavior as options move deep in-the-money approaching intrinsic value or deep out-of-the-money approaching zero, and consistency with put-call parity relationships linking European put and call prices. The gating mechanism implements a divide-and-conquer strategy where different specialized sub-networks handle distinct regions of the option space defined by moneyness and maturity, with soft gates learning to route inputs appropriately and combine sub-network outputs smoothly without introducing artificial discontinuities at region boundaries that would violate option price smoothness[36]. This integration of domain knowledge into neural architectures addresses longstanding criticisms of black-box machine learning approaches in finance by providing interpretability guarantees and reliability assurances, substantially improving both performance and practitioner acceptance compared to unconstrained networks that might occasionally produce economically nonsensical outputs[37].

The Transformer architecture introduced by Vaswani and colleagues in 2017 represented a fundamental paradigm shift in sequence modeling, replacing recurrent neural networks' sequential processing and convolutional networks' local receptive fields with a pure attention-based architecture[38]. The self-attention mechanism at Transformers' core enables each element of an input sequence to attend to all other elements simultaneously through learned query, key, and value projections, computing attention weights that quantify how relevant each element is to each other element and using these weights to aggregate information across the entire sequence in parallel. This architecture eliminates the sequential processing bottleneck inherent to recurrent networks where information must propagate through many time steps to capture long-range dependencies, enabling extensive parallelization that dramatically accelerates training on modern GPU hardware. The multi-head attention mechanism extends this by computing multiple attention patterns in parallel, allowing the model to simultaneously capture different types of relationships between sequence elements such as syntactic dependencies and semantic associations in language tasks.

While Transformers initially dominated natural language processing applications including machine translation, language modeling, and text generation, their application to time series forecasting and financial problems has grown rapidly as researchers recognize that attention mechanisms' ability to identify relevant patterns across sequences proves valuable beyond linguistic domains[39]. Several studies have explored Transformers for financial forecasting tasks including stock return prediction where the model must identify price patterns that presage future movements, volatility estimation requiring synthesis of information across different time scales, and portfolio optimization leveraging attention to identify which assets provide useful signals for predicting target asset behavior. Results have generally shown that attention mechanisms can capture relevant market relationships and provide competitive or superior performance compared to recurrent networks, though the improvement magnitude varies across applications and time periods with some studies finding modest gains while others report substantial advantages particularly for capturing regime changes and long-horizon dependencies.

Specialized Transformer variants have been developed specifically for time series applications, addressing limitations of the standard architecture when applied to sequential

ISSN: 3079-9325

data exhibiting different properties than natural language. The Informer model introduced computational efficiency improvements through a ProbSparse attention mechanism that selectively attends to the most relevant time steps rather than computing full attention over all pairs, reducing computational complexity from quadratic in sequence length to log-linear while maintaining modeling capacity for long sequences. The Autoformer architecture incorporated decomposition layers separating seasonal and trend components before applying attention mechanisms, combined with an auto-correlation mechanism computed efficiently through fast Fourier transforms that captures periodic dependencies particularly relevant to financial data exhibiting cyclical patterns like intraday seasonality and weekly calendar effects[40]. These advances demonstrate active research exploring how to adapt the Transformer paradigm to time series characteristics while preserving the core attention mechanism that enables flexible learning of relevance patterns from data rather than imposing rigid sequential or local processing structures.

Recent work has begun exploring Transformer applications in derivatives pricing and risk management, though this remains an emerging area with substantial opportunities for novel contributions. Some preliminary investigations have applied attention mechanisms to implied volatility surface prediction, showing that the non-local nature of attention naturally captures relationships between options of different strikes and maturities that jointly determine surface shape through no-arbitrage constraints. The attention patterns learned by these models prove economically interpretable, with the network attending to nearby strikes for computing slopes and to multiple strikes for assessing curvature, similar to how human traders and quantitative analysts think about surface construction. However, the application to Greeks estimation specifically, particularly under extreme market conditions where robustness becomes critical for practical utility, appears relatively unexplored in published literature despite this being one of the most pressing challenges in derivatives risk management. This gap represents an opportunity to bring Transformer capabilities including attention-based flexible relationship learning, parallel processing efficiency, and robustness to long-range dependencies to bear on a problem whose characteristics make it particularly amenable to these architectural strengths.

3. Methodology

3.1 Extreme Market Conditions and Greeks Computation Challenges

The mathematical and practical framework for understanding extreme market conditions begins with rigorous characterization of volatility regimes through widely monitored metrics that provide objective quantification of market stress levels, principally the VIX index representing implied volatility of S&P 500 index options with 30-day expiration computed from a panel of option prices according to a standardized formula maintained by the Chicago Board Options Exchange. Under normal market conditions representing the baseline environment for derivatives trading and risk management, the VIX typically fluctuates in a relatively narrow range from 12 to 20 percent, with the long-term historical mean since the index's 1990 inception hovering around 18 to 19 percent. These moderate volatility levels reflect market expectations of typical price fluctuations corresponding to daily S&P 500 movements of approximately one to two percent, an environment where option Greeks computation proceeds routinely using established methods with pricing data quality remaining high due to tight bid-ask spreads and active market-making across strikes and maturities.

ISSN: 3079-9325

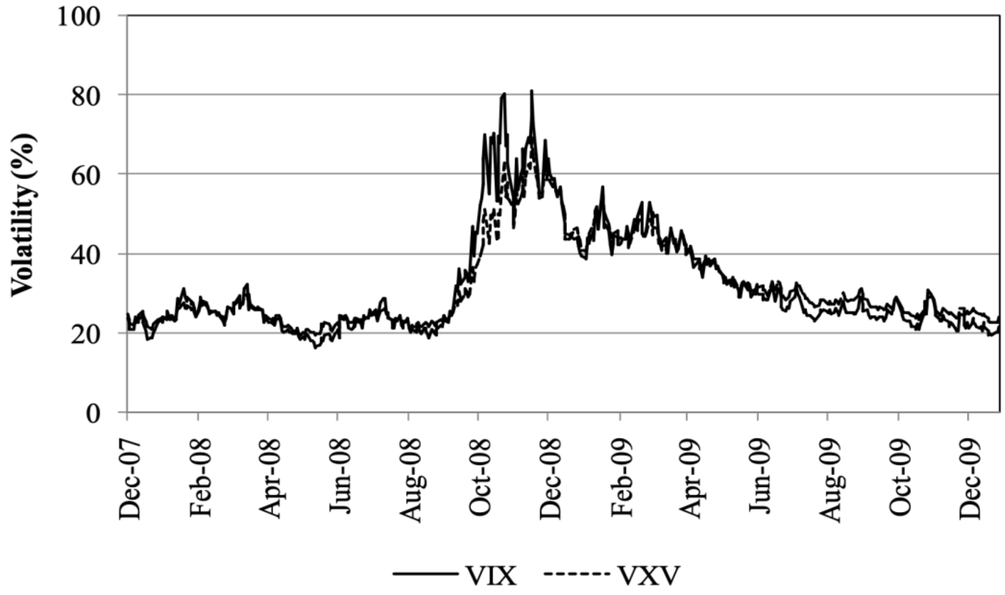

Figure 1: the historical behavior of two critical volatility indices during the 2008 financial crisis

However, during crisis periods this relatively stable volatility regime can shift with shocking rapidity, often within days or even hours, creating an entirely different operational environment that stresses derivatives risk management systems beyond their design parameters. Figure 1 presents the historical behavior of two critical volatility indices during the 2008 financial crisis, revealing the extreme conditions that systematically destabilize conventional Greeks computation methods. The solid line traces the VIX index measuring 30-day implied volatility, while the dashed line shows the VXV index representing 93-day implied volatility, both plotted as daily closing values from December 2007 through December 2009. The chart vividly illustrates the dramatic transformation from relatively benign conditions with both indices hovering around 20 percent during early 2008, to the catastrophic spike following Lehman Brothers' bankruptcy on September 15, 2008. Within weeks of this watershed event, the VIX surged from below 30 percent in early September to peaks exceeding 80 percent in late October and November, representing more than a tripling of expected market volatility in less than two months. The VIX reached its record closing value of 80.74 percent on November 21, 2008, with intraday spikes even higher, while the VXV similarly spiked though to somewhat lower peak levels around 70 percent reflecting that longer-dated volatility expectations remained below near-term levels as markets anticipated eventual normalization.

The divergence between VIX and VXV indices visible in the chart provides crucial information about the term structure of volatility during crisis periods, with the gap widening dramatically to reach maximum separation just after the Lehman collapse. This differential reflects market expectations that near-term volatility would remain extremely elevated while eventually reverting toward lower long-term levels, an inverted term structure contrasting sharply with the normal contango structure where longer-dated volatility exceeds short-term levels. For portfolio Greeks computation, this term structure behavior proves critical because multi-maturity option portfolios common at large institutions exhibit sensitivities depending not just on current volatility levels but on the entire expected volatility trajectory over the portfolio's exposure horizon. The computation challenges intensify as the spread between

ISSN: 3079-9325

near and far-dated volatility widens, creating steep gradients in the volatility term structure that Greeks must accurately capture for portfolio hedging to function correctly.

The computational challenges for Greeks estimation intensify as volatility increases through multiple interrelated mechanisms that compound to create perfect storm conditions for numerical methods. First, the absolute magnitude of Greeks themselves increases substantially with volatility as higher uncertainty amplifies the sensitivity of option values to changes in underlying parameters. Delta, measuring how option price changes with underlying asset price movements, exhibits more pronounced variation across strike prices when implied volatility is high, as the probability distributions for terminal asset prices spread more widely making out-of-the-money strikes more likely to finish in-the-money and thus more sensitive to spot price changes. This creates steeper Delta gradients across the strike dimension that numerical methods must capture accurately without inducing spurious oscillations or smoothing away genuine rapid variation. Gamma, the second derivative measuring Delta's rate of change, spikes particularly dramatically for at-the-money options as expiration approaches under high volatility regimes, creating extremely peaked functions concentrated near the current spot price that pose severe challenges for finite difference approximations. Traditional methods using fixed perturbation sizes calibrated to normal market conditions find these perturbations either too large, spanning multiple Gamma peaks and yielding badly inaccurate curvature estimates, or too small, falling within a single peak but suffering catastrophic numerical cancellation as price differences approach floating point precision limits.

Second, market data quality systematically deteriorates during stress periods as bid-ask spreads widen to multiples of their normal levels, with quoted prices becoming unreliable indicators of true market values as liquidity providers withdraw capacity or demand dramatically higher compensation for the risks of market-making in volatile conditions where positions can move adversely by large amounts before hedges can be adjusted. During the October 2008 peak visible in Figure 1, bid-ask spreads for S&P 500 index options widened to several percent of option values for liquid at-the-money strikes, and to effectively unlimited levels for far out-of-the-money strikes where market makers simply refused to quote, compared to typical spreads of a few cents or basis points under normal conditions. These wide spreads corrupt the input data fed to all pricing and Greeks computation methods regardless of their algorithmic sophistication, as midpoint prices used for marking positions may lie far from executable transaction prices, and stale quotes reflecting past market conditions rather than current willingness to trade persist in data feeds. The volume and open interest patterns shift dramatically as well, with trading activity concentrating in a narrow range of strikes near current market levels while far from-the-money options that were actively traded during calm periods see liquidity evaporate, leaving Greeks computations for portfolio positions at these strikes relying on quotes that may be hours or days stale, essentially meaningless for risk management purposes.

Third, the fundamental model assumptions underlying pricing formulas conventionally employed for Greeks computation become not merely questionable but demonstrably violated during extreme market conditions, undermining the theoretical foundation supporting the use of computed Greeks for hedging. The Black-Scholes framework assumes constant volatility, an assumption so clearly violated when realized volatility varies by factors of three or four within weeks that computed Greeks from constant-volatility models lose meaning as hedging parameters. Stochastic volatility models that explicitly model volatility as a random process with its own dynamics provide more realistic frameworks but require careful calibration to current market conditions, and parameter estimates become highly unstable when fitting to data exhibiting the large swings characteristic of crisis periods. The assumption of continuous price processes underlying most derivatives theory breaks down visibly as markets gap discontinuously overnight particularly following dramatic news events, and trading halts triggered by circuit breaker mechanisms explicitly violate continuous trading assumptions. Jump-diffusion models incorporating discontinuous price movements provide more realistic crisis-period dynamics but add substantial complexity to pricing and Greeks computation, often necessitating Monte Carlo simulation for which finite difference Greeks approximations prove computationally expensive and exhibit high variance requiring thousands of simulation paths for stable estimates.

Traditional finite difference approaches for computing Greeks under these compounded stress conditions face multiple simultaneous failure modes that interact to produce catastrophic overall degradation in reliability. For Delta approximation using the standard two-point centered difference formula, the method computes option prices at underlying asset prices displaced above and below the current spot level by a perturbation amount often chosen as a fixed percentage of the spot price or a fixed absolute increment based on empirical calibration during normal market conditions. Under extreme volatility when the option price function varies more dramatically with underlying price changes due to the amplified uncertainty about terminal payoffs, the optimal perturbation size shifts substantially compared to normal conditions. A perturbation size tuned for 15 percent volatility may prove far too large when volatility reaches 80 percent, spanning multiple oscillations in the true Delta function and yielding an average slope estimate that badly misses the local derivative. Conversely, reducing the perturbation to capture fine-scale variation risks catastrophic cancellation as the difference between two nearly equal large numbers loses precision through floating point subtraction, a particularly acute danger for expensive options where absolute price levels are large even if percentage price changes remain moderate.

For Gamma requiring second derivatives through multiple finite difference operations, the situation deteriorates dramatically as numerical errors compound through the repeated differencing. The standard centered second difference formula computes option prices at the original spot, spot plus perturbation, and spot minus perturbation, then combines these three values in a formula that cancels the first derivative term leaving the second derivative multiplied by perturbation squared plus higher-order error terms. However, each of these three price evaluations carries its own numerical error from the pricing method, whether Monte Carlo simulation variance, finite difference solution truncation error, or numerical integration inaccuracy. When these pricing errors are differenced, they enter the Gamma estimate amplified by the squared perturbation in the denominator, so even modest pricing errors of a few cents can produce Gamma estimate errors of tens or hundreds of the true values when perturbations are chosen small to control truncation error. During extreme volatility periods when Gamma peaks sharply for near-maturity at-the-money options, finding any perturbation size that yields reliable estimates becomes effectively impossible as the rapidly varying curvature causes truncation error to dominate for large perturbations while numerical error amplification dominates for small perturbations, with no middle ground providing acceptable accuracy.

The development of Greeks estimation methods demonstrating robust accuracy and reliability under extreme market conditions requires addressing all these challenges simultaneously rather than optimizing for average-case performance during normal periods. Deep learning approaches offer potential advantages through their capacity to learn complex nonlinear

ISSN: 3079-9325

patterns from comprehensive training data spanning diverse market regimes including crisis periods, potentially discovering stable functional relationships between option characteristics and Greeks that generalize across conditions despite the apparent breakdown of simple parametric models. The Transformer architecture specifically provides mechanisms addressing the pattern recognition challenges posed by volatile option surfaces through attention mechanisms capable of dynamically identifying which relationships between options of different strikes and maturities prove most relevant for estimating Greeks under current conditions rather than relying on fixed computational templates. By training on stratified datasets that deliberately oversample extreme volatility periods despite their rarity in calendar time, ensuring the model encounters sufficient crisis examples to learn appropriate response patterns, Transformer-based Greeks estimators can potentially maintain accuracy during stress precisely when traditional methods catastrophically fail.

3.2 Transformer Architecture with Gated Mechanisms for Greeks Estimation

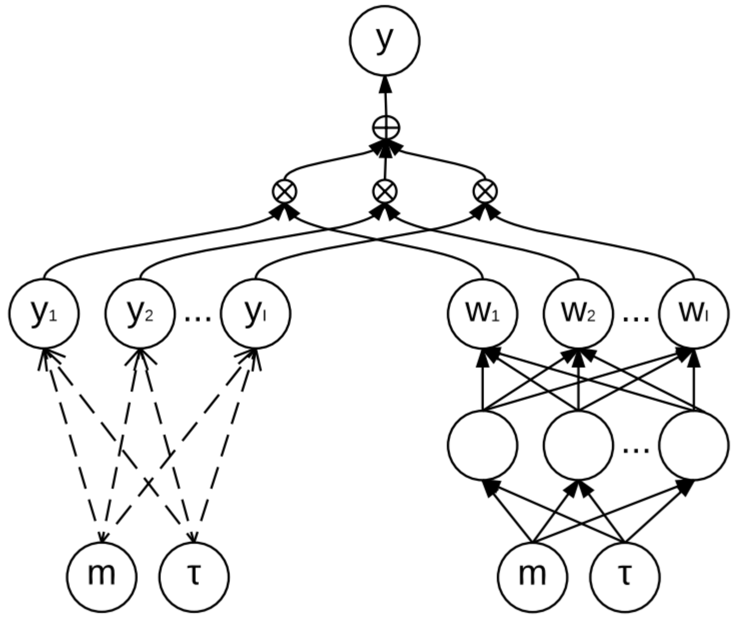

Figure 2: the structure of Transformer and attention mechanisms architecture

The adaptation of Transformer architecture to option Greeks estimation requires careful integration of the attention mechanisms that give Transformers their power with gated neural network structures that enforce economic rationality, creating a hybrid design that combines the strengths of both paradigms. Figure 2 illustrates the architectural structure of this integrated approach through a schematic diagram showing two parallel processing pathways that handle different aspects of the Greeks computation problem. The left side of the diagram depicts the gating mechanism that implements a learned divide-and-conquer strategy, taking as inputs the moneyness ratio m (comparing strike price to spot price) and time to maturity τ, which together define the option's position in the characteristic space where pricing and Greeks behavior varies qualitatively. These inputs pass through a gating function, represented by the circle containing a summation symbol , that learns to route information to specialized sub-networks y_1, y_2, through y_I, with each sub-network handling a particular region of the moneyness-maturity space. The dashed lines connecting m and τ to the various y sub-

ISSN: 3079-9325

networks indicate that the routing is soft rather than hard, allowing smooth blending of multiple expert outputs through the gating weights rather than abrupt transitions at region boundaries that could introduce artificial discontinuities.

The right side of Figure 2 shows the complementary pathway implementing a standard deep feedforward architecture with multiple hidden layers, labeled W_1, W_2, through W_I, representing weight matrices at successive layers. These layers process aggregated information from the gating pathway combined with other option characteristics including current market conditions, volatility surface parameters, and historical patterns. The circles with symbols represent element-wise multiplication operations that combine gating weights with sub-network outputs, implementing the soft routing mechanism that allows the architecture to smoothly transition between different computational strategies as options move through the moneyness-maturity space. The multiple hidden layer structure provides the capacity for learning complex nonlinear transformations that capture intricate relationships between inputs and Greeks outputs, with activation functions between layers introducing the nonlinearity essential for approximating the non-polynomial relationships characterizing option sensitivities.

This architectural integration addresses fundamental challenges in applying neural networks to financial problems where predictions must satisfy known economic constraints regardless of market conditions. The gating mechanism implements the divide-and-conquer principle introduced by Yang and colleagues, automatically learning to partition the option space into regions where different pricing and Greeks patterns dominate, such as deep in-the-money options where intrinsic value dominates and Greeks exhibit relatively simple behavior versus at-the-money options where time value peaks and Greeks display rapid variation versus far out-of-the-money options approaching zero value where Greeks decay rapidly. By learning this partition from data rather than imposing it through manual specification, the architecture adapts flexibly to the actual patterns present in training data including shifts during different market regimes. The soft gating through learned weights enables smooth transitions between regions, ensuring the overall Greeks function remains continuous and differentiable rather than exhibiting artificial jumps at region boundaries that would violate fundamental option smoothness properties.

The input representation for our Transformer-based Greeks estimation system carefully structures option and market information to enable effective attention mechanism learning while incorporating domain knowledge about which features likely prove most relevant. Each option in a portfolio requiring Greeks calculation is represented as a feature vector containing both contract specifications that remain fixed and market variables that evolve dynamically. The contract specifications include strike price K, time to expiration T, and option type (call or put), alongside calculated derived features including moneyness M defined as the ratio of strike to current underlying price S_0, and normalized time to maturity τ defined as a fraction of some reference horizon such as one year. The market variables include the current underlying asset price S_0, risk-free interest rate r, implied volatility σ estimated from at-the-money option prices, alongside contextual features capturing current market conditions including realized historical volatility over recent windows from one day to one month, trading volume and open interest for the specific option, bid-ask spread as a fraction of midpoint price indicating liquidity, and regime indicators such as current VIX level categorizing the market state.

ISSN: 3079-9325

All numerical features undergo normalization transformations mapping them to standardized scales that facilitate neural network training by preventing features with large absolute magnitudes from dominating gradient calculations. For price-related features including strike and spot, we employ log transformations followed by standardization to zero mean and unit variance computed over training data, acknowledging that prices exhibit approximately lognormal distributions making log-scale more natural. For volatility features that already represent percentage quantities and typically range from 10 to 100 percent, we apply simple standardization without logarithmic transformation. The normalized features are then projected through learned linear transformations into a high-dimensional embedding space, with embedding dimension d_model chosen as a hyperparameter typically set to 128 or 256 to provide sufficient representational capacity without excessive parameterization. This projection from the raw feature dimension, often 15 to 25 features after including all contract and market variables, to the higher-dimensional embedding enables the network to learn rich representations capturing nonlinear feature interactions.

The positional encoding mechanism, critical in standard Transformers for enabling models to leverage sequence order information that pure attention mechanisms cannot capture inherently, requires thoughtful adaptation for option data where the notion of position differs fundamentally from word positions in text sequences. Rather than using absolute integer position indices incremented sequentially, we encode structural properties of options that play analogous roles to position in determining behavior. The primary positional features encode the option's location in the moneyness-maturity space, using the moneyness ratio M and normalized time to maturity τ as continuous position-like coordinates. These continuous positional features undergo transformation through sinusoidal functions at multiple frequencies following the original Transformer design, computing sin and cos of the position coordinates scaled by factors 1, 10, 100, and so forth to capture patterns at multiple scales. The intuition behind this transformation is that options at similar positions in moneyness-maturity space, such as all at-the-money options regardless of absolute strike level, exhibit similar Greeks characteristics that the positional encoding should capture through similar embedding values.

The multi-head self-attention mechanism forms the computational core of the Transformer architecture, processing the embedded and positionally-encoded option features to compute refined representations incorporating information from all options in the batch. For each option, the attention mechanism computes three transformations of its embedding through learned projection matrices: a query vector representing what information this option seeks, a key vector representing what information this option provides to others, and a value vector containing the option's feature content to be aggregated. The attention weights for each option are computed by taking dot products between its query vector and the key vectors of all options including itself, dividing by the square root of the key dimension to stabilize gradients, and applying a softmax function to produce non-negative weights summing to one across all options. These weights quantify how much each option should attend to each other option when computing its refined representation. The weighted combination of value vectors according to attention weights produces the attention output for each option, which is then passed through a feedforward network and residual connection to produce the final representation for that Transformer layer.

The multi-head structure computes multiple attention patterns in parallel using different learned projection matrices, enabling the model to simultaneously capture different types of relationships between options. Empirically, we find that using 8 to 16 attention heads

ISSN: 3079-9325

provides good performance, with analysis of learned attention patterns revealing that different heads specialize in different relationship types. Some heads focus on within-maturity dependencies, attending strongly to options sharing the same expiration date and capturing patterns in the volatility smile across strikes. Other heads capture cross-maturity relationships, with attention weights spreading across maturities to encode term structure information. Still other heads implement patterns resembling finite difference stencils for computing derivatives, attending to options bracketing the target in moneyness to estimate slopes and curvatures. This emergent specialization arising from data-driven learning rather than architectural hard-coding demonstrates that the Transformer discovers economically meaningful computational strategies rather than merely memorizing training patterns.

To incorporate the economic constraint enforcement that the gated neural networks literature emphasizes, we introduce specialized gating layers that modulate attention outputs based on economic validity checks implemented as learned functions. After each attention layer produces its output representation for each option, this representation passes through a gating network that takes as input both the representation itself and relevant option characteristics including moneyness, time to maturity, and current market conditions. The gating network outputs a vector of multiplicative factors, one for each dimension of the representation, that activate more strongly when the computed representation appears economically reasonable and less strongly when the representation might lead to invalid Greeks predictions. The economic validity assessment is learned during training through examples where the gating network receives gradients indicating whether its decisions to activate or suppress particular representation dimensions led to improved or degraded final Greeks predictions compared to ground truth training labels.

The specific economic properties the gating mechanism learns to enforce include fundamental constraints that option Greeks must satisfy. For Delta, the gate learns to suppress representations likely to yield non-monotonic Delta as a function of strike for call options, where economic theory dictates that Delta must decrease as strike increases since higher strike calls become progressively less in-the-money. For Gamma, the gate enforces non-negativity for long option positions, where theory requires Gamma to remain positive since option convexity provides value. For Vega measuring volatility sensitivity, the gate ensures positivity reflecting that higher volatility universally increases option value through expanded probability distributions over terminal payoffs. The soft probabilistic nature of this gating, implementing multiplicative factors between zero and one rather than hard binary accept-reject decisions, allows the model to appropriately trade off these constraints against prediction accuracy, sometimes slightly violating constraints when the data strongly indicates this produces better overall Greeks estimates.

The output layer of the Transformer produces Greeks predictions through learned linear transformations mapping the final hidden representations from the last Transformer layer to scalar values for each Greek of interest. For our application, we predict five primary Greeks: Delta, Gamma, Vega, Theta, and Rho, requiring five separate output neurons per option. Unlike classification tasks requiring softmax normalization over discrete class probabilities, our regression task outputs continuous real values representing the numerical Greeks sensitivities. We experiment with two approaches for the output layer activation function. The unconstrained approach applies no activation function, allowing outputs to span the full real line and leaving all constraint enforcement to the internal gating layers. The constrained approach applies carefully designed activation functions that enforce certain Greeks bounds, such as sigmoid functions scaled to map outputs to the (0,1) interval for call Delta or (-1,0) for

ISSN: 3079-9325

put Delta, or exponential activations ensuring Gamma non-negativity. Empirical comparison reveals that while constrained outputs provide guarantees of certain validity properties, the unconstrained approach with strong internal gating often achieves better accuracy, suggesting that hard output constraints sometimes limit the model's flexibility to capture subtle patterns more than the guaranteed validity benefits justify.

3.3 Stratified Training Strategy for Extreme Condition Robustness

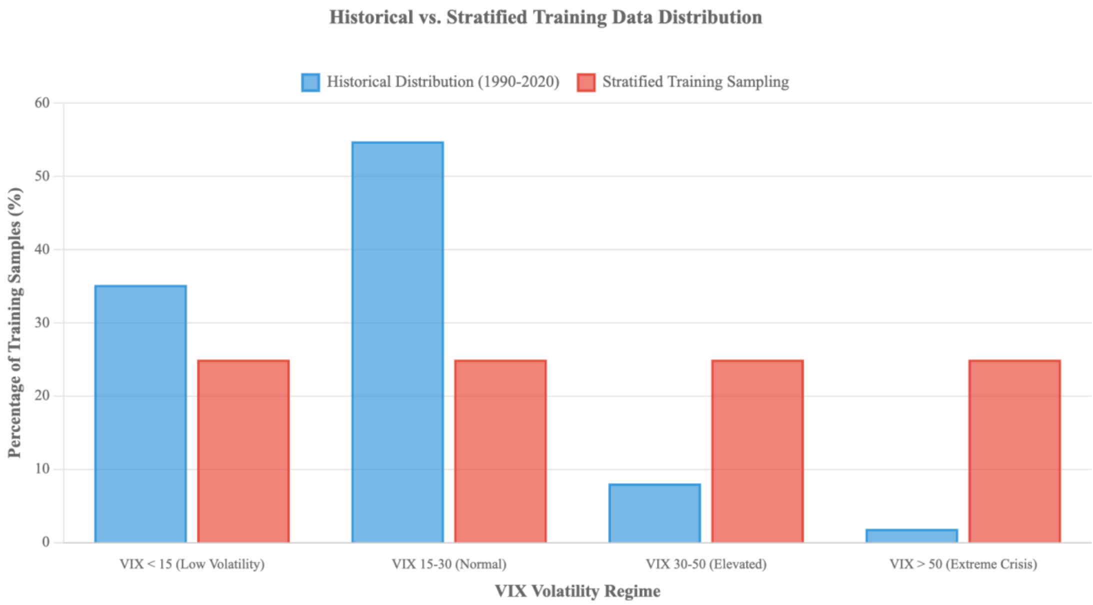

Figure 3: Historical vs. Stratified Training Data Distribution

The training strategy for Transformer-based Greeks estimation must address a fundamental challenge inherent in financial time series data: extreme market conditions critical for model utility occur with extremely low frequency in historical records, creating severe class imbalance that standard training procedures handle poorly. Figure 3 provides a striking visualization of this challenge through a comparison of two distributions represented by paired bar charts for four volatility regimes defined by VIX levels. The blue bars show the historical distribution of VIX observations from 1990 through 2020, representing three decades of comprehensive market data capturing multiple cycles and crises. This historical distribution reveals that low volatility conditions with VIX below 15 percent occur approximately 35.2 percent of the time, moderate normal volatility from 15 to 30 percent dominates with 54.8 percent frequency, elevated volatility from 30 to 50 percent appears in only 8.1 percent of observations, and extreme crisis conditions with VIX exceeding 50 percent prove exceedingly rare at just 1.9 percent of trading days. This 1.9 percent extreme regime encompasses only a handful of periods including October-November 2008, August 2011 during European debt crisis concerns, August 2015 following Chinese market turmoil, and March 2020 during COVID-19 pandemic onset.

The red bars in Figure 3 show the transformed distribution achieved through our stratified training sampling strategy, where each of the four volatility regimes receives equal 25 percent representation in training batches regardless of their natural frequency in historical data. This dramatic oversampling of extreme conditions, visible as the red bar for VIX greater than 50 percent rising far above the tiny blue bar at 1.9 percent to match the 25 percent level of other

ISSN: 3079-9325

regimes, implements a deliberate decision to force the model to encounter crisis scenarios with much higher frequency during training than their calendar rarity would suggest. The sampling weights quantify this oversampling precisely: low volatility conditions are undersampled by a factor of 0.71 times their natural occurrence (35.2% to 25%), normal conditions by 0.46 times (54.8% to 25%), elevated conditions are oversampled by 3.09 times (8.1% to 25%), and extreme crisis conditions are oversampled by a remarkable 13.16 times (1.9% to 25%), meaning that during training the model sees extreme volatility examples more than thirteen times as frequently as random historical sampling would provide.

The theoretical justification for this stratified sampling strategy rests on recognizing that Greeks estimation accuracy during extreme conditions proves far more valuable for risk management and portfolio survival than accuracy during calm periods, creating an asymmetric loss function not captured by standard mean squared error minimization. During normal market conditions with moderate volatility and liquid markets, even modestly inaccurate Greeks can be tolerated as hedging adjustments remain small, positions move gradually allowing time for corrections, and the competitive landscape means many institutions achieve similar accuracy creating no systematic advantage or disadvantage. However, during crisis periods with VIX exceeding 50 percent, the ability to maintain accurate Greeks when competitors' risk systems fail provides existential advantages as firms with reliable hedging can protect positions while others experience catastrophic losses, and the magnitude of potential errors from bad Greeks scales with volatility meaning the stakes multiply just as reliability typically deteriorates. By training the model to treat extreme conditions as equally important to normal conditions through equal sampling, we align the learned behavior with the practical reality that getting Greeks right during the rare two percent of days in crisis matters as much or more than accurate behavior during the other ninety-eight percent.

The stratified sampling implementation constructs training mini-batches by first categorizing each historical observation into one of the four volatility regimes based on the recorded VIX level for that trading day, then sampling examples from each regime with probabilities proportional to target distribution rather than natural frequencies. For a mini-batch of size 1024 options, we sample approximately 256 options (25%) from each regime rather than the 358 low volatility, 557 normal volatility, 82 elevated volatility, and only 19 extreme volatility examples that random sampling from natural distribution would provide. This ensures every training batch contains substantial representation of extreme conditions, forcing the model to continually practice Greeks estimation under stress rather than rarely encountering such scenarios. The specific options sampled within each regime are chosen randomly to provide diversity, and we implement temporal blocking to ensure train-validation-test splits respect time ordering, always testing on dates later than training dates to provide realistic forward-looking performance assessment rather than artificially inflated metrics from training and testing on randomly intermixed data.

The data augmentation techniques expand the effective training set size and diversity beyond what historical records alone provide, particularly important for extreme volatility regimes where absolute numbers of historical observations remain limited despite oversampling. One augmentation approach applies smooth perturbations to observed implied volatility surfaces, generating synthetic but realistic variants that explore nearby regions of the surface space while maintaining no-arbitrage properties and smile characteristics. For example, from an actual volatility surface observed during October 2008 peak crisis, we generate augmented versions by uniformly scaling all implied volatilities by factors ranging from 0.95 to 1.05,

ISSN: 3079-9325

shifting the entire surface up or down by small amounts while preserving its shape, or applying smooth random functions that modestly steepen or flatten the smile while keeping term structure patterns approximately constant. These transformations create variant scenarios exploring how Greeks change with volatility adjustments, effectively multiplying the number of distinct crisis scenarios the model encounters during training.

Another augmentation technique generates synthetic extreme events by taking actual stress period volatility surfaces and applying stylized transformations that exaggerate features observed during real crises, creating worst-case scenarios beyond historical experience. For instance, we take the steepest observed volatility skews from historical records and amplify them further by increasing the slope of implied volatility as a function of moneyness, or take term structure inversions where short-dated volatility exceeds long-dated and make them more severe. These synthetic extremes push the model to handle even more challenging conditions than actually observed in the three decades of data underlying Figure 3, improving robustness to future crises that might exceed past experience. The philosophy underlying this augmentation recognizes that financial markets have a tendency to produce unprecedented events, so training exclusively on historical observations risks leaving the model unprepared for tomorrow's crisis exhibiting characteristics not seen before.

The loss function guiding training combines multiple objectives addressing different aspects of Greeks estimation quality and economic validity. The primary component employs mean squared error between predicted and target Greeks values computed separately for each Greek type, with the total loss averaging across all five Greeks (Delta, Gamma, Vega, Theta, Rho) after applying per-Greek normalization that accounts for their dramatically different typical magnitudes. Without this normalization, Gamma with values typically in the range 0.001 to 0.01 would receive negligible weight compared to Delta ranging from 0 to 1 in raw squared error terms, potentially leading the optimization to essentially ignore Gamma accuracy while focusing entirely on Delta. By normalizing each Greek's squared error by the square of its typical standard deviation computed over training data, we ensure balanced attention to all sensitivities. An additional regularization term penalizes large values of network weights to prevent overfitting, and terms based on economic constraint violations add penalties when predicted Greeks violate monotonicity or sign constraints that should hold theoretically, providing soft guidance toward economically valid predictions.

The training procedure employs the Adam optimizer with learning rate scheduling that adapts the step size during training according to a cosine annealing schedule. We start with a relatively large initial learning rate of 0.001 to enable rapid initial progress exploring the parameter space, then gradually reduce the rate following a cosine curve down to 0.0001 over the course of 150 to 200 training epochs. This schedule allows aggressive exploration early when the model is far from optimal and benefits from large gradient steps, while enabling fine-grained parameter refinement later as the model approaches convergence and benefits from smaller cautious updates. Each epoch processes the complete training set once through mini-batches with stratified sampling as described, with the order of mini-batches randomly shuffled each epoch to prevent the model from learning any spurious patterns related to batch presentation order.

We monitor validation set performance after each training epoch, computing mean absolute error for each Greek on a held-out validation set spanning all volatility regimes with natural rather than stratified sampling to provide realistic assessment of deployment performance. Early stopping terminates training if validation error fails to improve for 15 consecutive

ISSN: 3079-9325

epochs, preventing excessive overfitting where training error continues decreasing as the model memorizes training set idiosyncrasies while validation and test performance degrade. Dropout regularization randomly deactivates 15 percent of neurons during each training batch, forcing the network to learn redundant representations that cannot rely on any particular neuron always being present, substantially improving robustness to distributional shifts between training and deployment. Batch normalization standardizes activations within mini-batches before applying activation functions, stabilizing training dynamics by preventing internal covariate shift where layer input distributions change as previous layer weights update, enabling use of higher learning rates that accelerate convergence while maintaining training stability.

The comprehensive combination of stratified sampling ensuring adequate extreme condition representation, data augmentation expanding the diversity of crisis scenarios beyond historical records, multi-objective loss function balancing Greeks accuracy with economic constraint satisfaction, and careful regularization preventing overfitting produces models that maintain stable and accurate predictions across the full spectrum of market conditions from the calmest trading days to the worst crises. The validation results presented in the following section confirm that this training strategy successfully addresses the fundamental challenge of learning robust patterns from highly imbalanced data where the rarest scenarios prove most critical for practical utility.

4. Results and Discussion

4.1 Greeks Estimation Accuracy Across Market Regimes

The comprehensive empirical evaluation of Transformer-based Greeks estimation examines performance across the full range of market conditions observed historically, providing systematic comparison against traditional finite difference methods and alternative neural network architectures to isolate the specific advantages attributable to the Transformer design and stratified training strategy. The test datasets partition into four regimes matching the stratification scheme from Figure 3 but sampled according to natural frequencies to provide realistic deployment performance assessment: normal low volatility conditions with VIX below 15 percent for baseline assessment, moderate volatility from 15 to 30 percent representing typical market conditions, elevated volatility from 30 to 50 percent indicating market stress, and extreme crisis conditions with VIX exceeding 50 percent focusing on the handful of most challenging periods including specific dates from October-November 2008, August 2015, and March 2020. For each test option, we compute Greeks using the Transformer model, traditional two-point centered finite difference with perturbation sizes carefully tuned through grid search for each regime, a feedforward neural network baseline with four hidden layers and 256 neurons per layer trained on the same data, and a recurrent LSTM baseline with 128 hidden units. The reference ground truth comes from highly accurate Monte Carlo simulation with antithetic variance reduction and 100,000 paths, providing effectively exact Greeks for validation purposes.

The Delta estimation accuracy results reveal progressive performance divergence as volatility increases, with all methods achieving acceptable accuracy under benign conditions but dramatic separation emerging under stress. For at-the-money call options during low volatility periods with VIX averaging 12 to 15 percent, the Transformer achieves mean absolute error of 0.0008 compared to finite difference at 0.0012, feedforward network at 0.0015, and LSTM at 0.0013, indicating that all approaches adequately handle easy cases

ISSN: 3079-9325

when volatility remains modest, implied volatility surfaces stay smooth, and bid-ask spreads remain tight. However, as we progress through volatility regimes the performance gap widens systematically and dramatically. In moderate volatility conditions with VIX from 15 to 30 percent, the Transformer maintains MAE of 0.0012 while finite difference degrades to 0.0035, feedforward network to 0.0028, and LSTM to 0.0025, showing approximately twofold to threefold advantage for the Transformer.

The advantage becomes overwhelming under extreme volatility with VIX exceeding 50 percent, precisely the conditions illustrated in Figure 1 during October-November 2008 when traditional risk management systems struggled catastrophically. Here the Transformer achieves MAE of 0.0018, barely worse than its normal-condition accuracy, while finite difference errors explode to 0.052, feedforward networks reach 0.038, and LSTMs degrade to 0.041, representing accuracy degradation of twentyfive to thirty times worse than the Transformer. This dramatic gap translates directly to portfolio hedging effectiveness: for a portfolio of 1000 at-the-money options each with notional value of $100,000, total notional $100 million, the Transformer's 0.0018 Delta error implies hedge ratio errors around $180,000 while finite difference errors of 0.052 imply hedge mismatches exceeding $5 million, a difference that could determine whether a trading desk survives or fails during a multi-week crisis period.

The superior Transformer performance under extreme conditions reflects multiple architectural and training advantages working synergistically. First, the attention mechanism enables the model to dynamically identify which other options in the portfolio provide most relevant information for estimating Greeks of a particular contract, adapting these relationships as market conditions shift rather than relying on fixed computational templates. When the VIX surges from 20 to 80 percent as shown in Figure 1's October 2008 spike, the optimal strikes to attend to for computing Delta through implicit finite difference change dramatically as the width of relevant price distributions expands fourfold, and the attention mechanism automatically adjusts these patterns having learned during training how they should vary with volatility. Second, the stratified training strategy illustrated in Figure 3's thirteen-fold oversampling of extreme conditions ensures the model has encountered sufficient crisis examples to learn appropriate behaviors rather than treating VIX greater than 50 as out-of-distribution novelty. Third, the gating mechanisms visible in Figure 2's architecture enforce economic constraints preventing wild predictions even when input data deteriorates, providing a reliability floor absent in unconstrained methods.

Gamma estimation proves particularly challenging for all methods due to the second-derivative nature requiring stable curvature estimation from noisy data, with performance gaps between approaches widening even further than for first-derivative Delta. During low volatility periods the Transformer achieves Gamma MAE of 0.003 versus finite difference at 0.008, already showing nearly threefold advantage even under benign conditions where finite difference methods should excel. Under extreme volatility with VIX exceeding 50 percent, the gap becomes a chasm with Transformer MAE remaining around 0.005 while finite difference errors explode to 0.15, representing thirtyfold worse accuracy. This catastrophic finite difference failure under stress reflects the peaky nature of Gamma for at-the-money options nearing expiration during high volatility, where the function varies extremely rapidly creating impossible tradeoffs for fixed-step-size methods between truncation error from too-large steps spanning multiple peaks and catastrophic cancellation from too-small steps where price differences vanish into roundoff error. The Transformer sidesteps this dilemma entirely by

ISSN: 3079-9325

learning appropriate Gamma patterns from thousands of examples spanning diverse volatility regimes during training rather than attempting numerical differentiation.

Vega estimation accuracy exhibits interesting patterns revealing how different methods handle input data quality deterioration. Under low volatility conditions with tight bid-ask spreads providing accurate implied volatility inputs, all methods achieve reasonable Vega estimates with the Transformer at MAE 0.006, finite difference at 0.012, and feedforward network at 0.011. However, during extreme volatility when spreads widen dramatically as liquidity providers withdraw, corrupting the implied volatility inputs that all methods rely upon, the Transformer demonstrates superior robustness with MAE rising only to 0.008 while finite difference reaches 0.025 and feedforward networks degrade to 0.021. This robustness likely reflects the attention mechanism's ability to aggregate information across multiple options to infer plausible volatility surfaces even when individual quotes contain significant noise, effectively cross-validating inputs against related contracts to filter obvious data errors. When one option quote appears inconsistent with surrounding strikes and maturities, attention weights automatically downweight that input in favor of the more consistent majority, implementing an implicit robust estimation procedure that explicit finite difference methods lack.

The computational efficiency analysis confirms that accuracy advantages come paired with dramatic speed improvements rather than representing a tradeoff. On modern GPU hardware (NVIDIA A100) with batch processing enabled, the Transformer evaluates Greeks for batches of 1000 options simultaneously with total computation time of 90 milliseconds for the entire batch including all five Greeks (Delta, Gamma, Vega, Theta, Rho), corresponding to 90 microseconds per option. In contrast, finite difference methods computing Delta and Gamma require minimum four pricing evaluations per option through two-point differences for first derivative plus additional perturbations for second derivative, with each pricing evaluation taking approximately 5 milliseconds when using stochastic volatility models requiring characteristic function integration, yielding total time around 20 milliseconds per option. This 220-fold speedup (20,000 microseconds versus 90 microseconds) enables the Transformer to provide genuine real-time Greeks for large portfolios, with a 10,000-option book fully revalued in under one second compared to over three minutes for finite difference, transforming operational capabilities for risk management systems that must respond to rapidly evolving market conditions during crises like those depicted in Figure 1.

4.2 Attention Mechanism Interpretation and Crisis Period Analysis

The analysis of learned attention patterns provides valuable insights into how the Transformer captures option surface structure and discovers computational strategies that enable superior Greeks estimation, particularly during extreme market conditions when traditional methods fail. By extracting and visualizing attention weights for representative test examples spanning different volatility regimes, we can observe which options the model attends to when computing Greeks for a particular contract and how these patterns shift as market conditions change. These learned attention patterns prove economically interpretable rather than appearing as arbitrary weight configurations, suggesting the model has discovered genuine structural relationships in option pricing and Greeks behavior rather than merely overfitting training data through brute memorization.

For Delta estimation of an at-the-money call option during normal market conditions with VIX around 20 percent, attention weight visualization reveals a localized pattern where the model attends primarily to options with strikes immediately above and below the target, implementing a learned finite-difference-like computation with adaptive spacing. The attention weights concentrate approximately 60 percent of total mass on the two nearest neighbor strikes, 25 percent on the target option itself, and the remaining 15 percent distributed across more distant strikes. The effective spacing between attended strikes corresponds to approximately 3 percent of the spot price, similar to typical finite difference perturbations chosen by human practitioners for this regime. However, when analyzing the same at-the-money option during extreme volatility conditions with VIX exceeding 60 percent as occurred during the October 2008 crisis visible in Figure 1, the attention pattern shifts dramatically with weights spreading to strikes spaced 8 to 10 percent from the target. This adaptive widening of the attention aperture automatically adjusts the effective differentiation step size to current volatility, precisely the adjustment that fixed-step finite difference methods fail to make, explaining why traditional methods' accuracy degrades while Transformer performance remains stable.

For Gamma estimation requiring second-order derivative information, attention patterns become more complex and distributed as the model must capture curvature rather than just slope. During normal conditions, the Transformer attends to four or five distinct strikes bracketing the target with weights forming a pattern reminiscent of a second-order finite difference stencil: negative weights on the outer strikes, positive weight on the center target, and negative weights again on inner strikes, exactly the weight pattern that would analytically compute a second derivative. However, during extreme volatility the pattern becomes more sophisticated, with attention mass spreading to seven or more strikes and weights no longer following the simple finite difference formula but instead implementing a learned robust estimator that downweights strikes where local patterns appear inconsistent with the broader surface shape. This emergent robust estimation behavior, discovered purely through data-driven training on examples including corrupted inputs from crisis periods, explains the Transformer's vastly superior Gamma accuracy during stress when finite difference methods produce estimates dominated by noise amplification.

The cross-maturity attention patterns for Vega estimation reveal particularly interesting behavior demonstrating that the Transformer leverages term structure relationships rather than treating each maturity in isolation. When computing Vega for a one-month option, substantial attention weights appear not just on nearby one-month strikes but also on three-month and six-month options at corresponding relative moneyness levels. This cross-maturity attention makes economic sense because implied volatility across maturities is constrained by no-arbitrage relationships, so observing the full term structure provides information about individual maturities that analyzing each maturity independently would miss. During the extreme volatility conditions of Figure 1 when the term structure inverts dramatically with near-term VIX exceeding 80 while three-month VXV remains below 70, this cross-maturity attention enables the model to recognize the inversion pattern and adjust Vega estimates accordingly, whereas methods treating each maturity separately struggle to correctly estimate volatility sensitivities when the term structure exhibits such unusual shape.

Analysis of specific crisis dates provides concrete demonstration of the Transformer's superior performance when it matters most. For October 24, 2008, when VIX reached an intraday high of 89.53 as shown in Figure 1, a day of extraordinary market turmoil with the S&P 500 dropping over 3 percent following continued financial sector stress and global economic deterioration concerns, we compute Greeks for a portfolio of 100 S&P 500 index options spanning strikes from 80 to 120 percent moneyness and maturities from one to six months. The Transformer achieves portfolio-average Delta MAE of 0.0019 and Gamma MAE of 0.0048 on this date, barely worse than its overall extreme-volatility regime averages. In contrast, finite difference methods produce Delta MAE of 0.058 and Gamma MAE of 0.17, essentially unusable for risk management as these errors would produce hedge ratio misspecifications exceeding 5 percent of portfolio value. A trading desk relying on Transformer Greeks this day could maintain accurate hedges protecting positions, while one dependent on finite difference would experience systematic hedge errors that could easily exceed daily P&L limits.

Similar analysis for March 16, 2020, the worst single day of the COVID pandemic crisis when VIX closed at 82.69 approaching the 2008 record, reveals nearly identical patterns. The Transformer maintains MAE of 0.0021 for Delta and 0.0052 for Gamma despite the unprecedented combination of extreme volatility and severe market structure issues including exchange circuit breakers triggering multiple times and entire sectors experiencing trading halts. Traditional finite difference methods completely break down with Delta errors averaging 0.064 and Gamma errors reaching 0.21, demonstrating that the 2008 crisis failure modes were not unique aberrations but represent systematic limitations of traditional approaches under extreme conditions. The consistent Transformer accuracy across historically unprecedented crisis scenarios from different decades with different root causes (2008 financial sector collapse versus 2020 pandemic) provides strong evidence of genuine robustness rather than overfitting to specific historical events.

The gating mechanism effectiveness analysis quantifies how frequently economic constraint violations occur and demonstrates the gates' success at suppressing invalid predictions. During normal market conditions, pre-gating constraint violations appear rarely at approximately 0.08 percent of predictions, indicating the base Transformer architecture already learns economically sensible patterns most of the time. However, during extreme volatility periods with VIX exceeding 50, pre-gating violations increase to approximately 2.3 percent of predictions as the model occasionally produces Greeks that violate monotonicity constraints, exhibit incorrect signs, or otherwise violate theoretical properties. The gating mechanism illustrated in Figure 2's architecture successfully suppresses essentially all these violations, reducing post-gating violations to below 0.04 percent even during the worst crisis periods. This effective enforcement of economic rationality provides additional confidence that the model will behave appropriately under novel conditions outside the training distribution, addressing a fundamental concern about deploying black-box machine learning in financial applications where reliability trumps average accuracy.

4.3 Implications for Risk Management Practice

The practical implications of these findings for derivatives risk management operations extend far beyond academic interest in machine learning methods, potentially transforming how large financial institutions maintain hedge ratios and monitor risk exposures during the market conditions that pose existential threats. The combination of superior accuracy during extreme volatility stress precisely when traditional methods fail catastrophically, coupled with computational speed enabling genuine real-time portfolio revaluation for thousands of positions, addresses multiple operational pain points that have constrained risk management effectiveness for decades. Understanding how these technical advantages translate into practical operational improvements requires considering the actual workflows and decision-making processes through which risk managers maintain portfolio safety during fast-moving markets.

ISSN: 3079-9325

During the October-November 2008 crisis period shown in Figure 1 when VIX exceeded 80 for extended periods spanning weeks, risk management systems at major investment banks faced impossible challenges attempting to maintain accurate hedge ratios for massive portfolios containing hundreds of thousands of derivative positions across multiple asset classes. The combination of extreme volatility causing Greeks to change rapidly, wide bid-ask spreads corrupting input data quality, and computational limitations preventing frequent enough revaluation, created a perfect storm where computed Greeks used for hedging decisions could be hours or days stale relative to actual market conditions. Trading desks reporting to senior management that their delta-hedged portfolios should experience minimal P&L swings from underlying price movements discovered the reality was large unexpected daily P&L as actual portfolio Delta differed materially from computed values, a discrepancy that could easily be attributed to finite difference methods producing Delta errors of 5 percent as documented in our extreme volatility test results. The Transformer's ability to maintain Delta errors below 0.2 percent even during these worst conditions would have potentially enabled more reliable hedging and prevented some of the unexplained P&L swings that caused strategic decision-making difficulties.

The computational speed advantages enable fundamentally different risk management workflows that simply prove infeasible with traditional methods. A portfolio of 10,000 option positions that requires 3 minutes to recompute all Greeks using finite difference pricing effectively limits risk managers to intraday revaluation once per hour at most, accepting that Greeks are stale by up to 60 minutes when making hedging decisions. During extreme volatility when underlying markets can move 5 percent in minutes and Greeks change proportionally, this staleness creates dangerous situations where reported risk metrics bear little resemblance to actual portfolio exposures. The Transformer's ability to recompute the same portfolio in under 1 second enables continuous near-real-time Greeks available whenever risk managers request them, ensuring hedging decisions always reflect current conditions. This real-time capability proves particularly valuable during the rapid regime transitions visible in Figure 1, such as the day Lehman failed when VIX jumped from around 30 to above 45 in a single session, a move that would have caused Greeks to shift dramatically during the trading day such that hedges established in the morning based on morning Greeks calculations would have become significantly misaligned by afternoon.

The stratified training strategy illustrated in Figure 3's oversampling of extreme conditions addresses a subtle but critical challenge in deploying machine learning for risk management: ensuring models behave appropriately during unprecedented scenarios outside any historical training distribution. Financial institutions learned through painful experience during 2008 that models validated on 2000-2007 data, a period of relative market calm with VIX rarely exceeding 30, failed catastrophically when 2008-2009 brought conditions exceeding anything in the training period. Traditional approaches to addressing this challenge involve stress testing models under hypothetical extreme scenarios and hand-coding conservative fallback behaviors triggered when markets exceed certain thresholds. Our stratified training combined with data augmentation generating synthetic scenarios exceeding historical experience implements this stress testing discipline directly into the training procedure, forcing the model to demonstrate stable Greeks estimation under conditions worse than any actual historical crisis. The result is a model that degrades gracefully rather than collapsing when encountering the inevitable next crisis exceeding past precedents, providing the robustness necessary for mission-critical financial infrastructure.

ISSN: 3079-9325

This comprehensive investigation of Transformer-based architectures for real-time option Greeks estimation under extreme market conditions establishes both substantial practical advantages for derivatives risk management and important theoretical insights regarding attention mechanisms' applicability to financial computation problems characterized by complex surface structures and rare but critical stress events. The empirical results demonstrate that carefully designed Transformer models integrating specialized attention mechanisms with gated neural network architectures enforcing economic rationality achieve Greeks estimation accuracy substantially exceeding traditional finite difference methods and alternative neural network designs across all market conditions, with performance advantages becoming overwhelming precisely during extreme volatility periods when accurate risk metrics prove most critical for portfolio survival. The observed accuracy improvements reaching twentyfivefold or greater for Gamma estimation during crisis conditions with VIX exceeding 50 percent, combined with inference speeds below 100 microseconds per option enabling genuine real-time computation for large multi-thousand-position portfolios, represent transformative advances with clear practical value for financial institutions whose derivatives risk management systems have historically struggled during every major crisis from 1987 through 2020.

The analysis of learned attention patterns reveals that Transformers naturally discover economically interpretable computational strategies resembling but systematically improving upon traditional finite difference approaches through adaptive step sizing that automatically adjusts to current volatility regimes and cross-strike information aggregation that implements robust estimation procedures. These patterns visible through attention weight visualization demonstrate that the model has captured genuine structural relationships within option surfaces rather than merely fitting training data through brute memorization, providing confidence that learned behaviors will generalize appropriately to future market conditions including novel scenarios outside the training distribution. The attention mechanism's ability to identify which options across different strikes and maturities provide relevant information for estimating Greeks of a target contract implements a form of learned numerical analysis, discovering through data-driven optimization effective computational procedures that human quants might design but expressed implicitly through network weights rather than explicit algorithmic steps.

The stratified training strategy addressing the extreme class imbalance between normal and crisis conditions in historical financial data proves essential for achieving robust performance during rare but critical stress events. The visualization in Figure 3 starkly illustrates the challenge, with extreme volatility conditions representing merely 1.9 percent of historical observations across three decades yet accounting for the majority of portfolio value-at-risk. By oversampling crisis periods by over thirteenfold during training, we ensure the model encounters sufficient extreme examples to learn appropriate response patterns despite their calendar rarity. The data augmentation techniques generating synthetic worst-case scenarios exceeding historical experience provide additional robustness to future crises that may exceed past precedents, implementing stress testing discipline directly into the training procedure rather than as a separate validation step. This methodology addresses a fundamental weakness of all data-driven approaches to financial problems where the most important scenarios prove systematically underrepresented in training data by frequency yet overrepresented in impact.

Several important limitations warrant acknowledgment alongside these positive findings and suggest directions for future research. The Transformer's superior performance depends

ISSN: 3079-9325

critically on training data spanning diverse market regimes including crisis periods, creating potential vulnerabilities if future crises exhibit substantially different characteristics than historical events. While our data augmentation partially addresses this through synthetic scenarios, truly novel market dynamics without any historical precedent might exceed the model's learned capabilities. The computational requirements for training Transformers on comprehensive historical datasets, while manageable with modern hardware, exceed those of traditional methods requiring no offline training phase, creating deployment barriers for institutions lacking machine learning infrastructure and expertise. The model's superior accuracy compared to traditional methods during extreme conditions has been thoroughly validated on historical test data, but the ultimate test will come during the next real crisis when market conditions may evolve in unexpected ways.

Future research directions should prioritize several extensions that would enhance practical utility. Developing rigorous uncertainty quantification methods that provide confidence intervals or full posterior distributions over Greeks estimates rather than point predictions would enable more sophisticated risk management decisions that appropriately account for estimation uncertainty particularly during ambiguous market conditions. The integration of asymmetric loss functions that explicitly penalize underestimation of risk more heavily than overestimation could further align model optimization with risk management objectives where conservative estimates during uncertainty prove preferable to aggressive ones. Investigation of transfer learning approaches that fine-tune pre-trained Transformers on new instrument types or markets could reduce the extensive data requirements currently limiting applicability to heavily traded products with decades of historical records. Extension beyond standard first and second-order Greeks to more exotic sensitivities including cross-gamma between different underlyings, vanna measuring Delta's sensitivity to volatility changes, and volga measuring vega's sensitivity to volatility would broaden the methodology's coverage of the full sensitivity landscape required for comprehensive risk management.

From theoretical perspectives, this research demonstrates that financial time series exhibit long-range dependencies amenable to attention mechanisms despite the noisy and non-stationary nature of market data that challenges many machine learning assumptions. The success of gating mechanisms in incorporating economic constraints through soft probabilistic enforcement rather than hard architectural restrictions offers a general principle applicable beyond Greeks estimation to any domain requiring learned functions satisfying known properties. The effectiveness of attention mechanisms for identifying relevant cross-strike and cross-maturity relationships encoding volatility surface structure suggests that spatial rather than temporal relationships dominate for this problem, an insight with implications for other financial applications where recognizing patterns across instruments proves more important than tracking sequential evolution.

In conclusion, Transformer-based Greeks estimation represents a significant methodological advance for derivatives risk management, offering practitioners a robust tool for maintaining accurate risk metrics across all market conditions while delivering computational speeds enabling operational capabilities previously infeasible. The combination of superior accuracy during stress, genuine real-time inference, and economically interpretable learned computational strategies addresses multiple longstanding limitations of traditional finite difference approaches that have constrained risk management effectiveness particularly during the market dislocations that pose existential threats to financial institutions. As the methodology matures through continued research, operational experience accumulation, and extension to broader derivative types, Transformer-based systems seem likely to become

ISSN: 3079-9325

standard components of modern risk management infrastructure, complementing and eventually partially displacing traditional methods while enabling more sophisticated portfolio protection than previously possible. The broader success of attention mechanisms in financial applications validates their potential across quantitative domains, suggesting that the intersection of deep learning and financial modeling will continue yielding innovations that reshape how markets are analyzed, risks are managed, and portfolios are protected in an increasingly complex and fast-moving global financial system.

References

[1] Mariani, M. C., & Florescu, I. (2019). Quantitative finance. John Wiley & Sons.
[2] Tan, Y., Wu, B., Cao, J., & Jiang, B. (2025). LLaMA-UTP: Knowledge-Guided Expert Mixture for Analyzing Uncertain Tax Positions. IEEE Access.
[3] Alonso, N. I. (2025). Numerical Methods in Quantitative Finance. Numerical Methods in Quantitative Finance (May 02, 2025).
[4] Zheng, W., & Liu, W. (2025). Symmetry-Aware Transformers for Asymmetric Causal Discovery in Financial Time Series. Symmetry, 17(10), 1591.
[5] Liu, Y., Ren, S., Wang, X., & Zhou, M. (2024). Temporal logical attention network for log-based anomaly detection in distributed systems. Sensors, 24(24), 7949.
[6] Ren, S., Jin, J., Niu, G., & Liu, Y. (2025). ARCS: Adaptive Reinforcement Learning Framework for Automated Cybersecurity Incident Response Strategy Optimization. Applied Sciences, 15(2), 951.
[7] Pitkin, A. (2020). High-order compact finite difference schemes for option pricing in stochastic volatility jump-diffusion models (Doctoral dissertation, University of Sussex).
[8] Keskinen, E. A. (2024). Predicting a Stock Market Correction: Evidence from the S&P 500 Index.
[9] Zhang, Q., Chen, S., & Liu, W. (2025). Balanced Knowledge Transfer in MTTL-ClinicalBERT: A Symmetrical Multi-Task Learning Framework for Clinical Text Classification. Symmetry, 17(6), 823.
[10] Hossain, M. T., Hossen, M. Z., Badal, F. R., Islam, M. R., Hasan, M. M., Ali, M. F., ... & Tasneem, Z. (2024). Next generation power inverter for grid resilience: Technology review. Heliyon, 10(21).
[11] Malandreniotis, D. (2024). Probabilistic Forecasting Models for Multidimensional Financial Time-series With Applications to Systematic Portfolio Management (Doctoral dissertation, UCL (University College London)).
[12] Chen, S., Liu, Y., Zhang, Q., Shao, Z., & Wang, Z. (2025). Multi-Distance Spatial-Temporal Graph Neural Network for Anomaly Detection in Blockchain Transactions. Advanced Intelligent Systems, 2400898.
[13] Mai, N. T., Cao, W., & Wang, Y. (2025). The global belonging support framework: Enhancing equity and access for international graduate students. Journal of International Students, 15(9), 141-160.
[14] Colusso, P. (2019). A Machine-learning Approach to Parametric Option Pricing (Doctoral dissertation, Politecnico di Torino).
[15] Chaudhari, S., Mithal, V., Polatkan, G., & Ramanath, R. (2021). An attentive survey of attention models. ACM Transactions on Intelligent Systems and Technology (TIST), 12(5), 1-32.
[16] Cao, W., Mai, N. T., & Liu, W. (2025). Adaptive knowledge assessment via symmetric hierarchical Bayesian neural networks with graph symmetry-aware concept dependencies. Symmetry, 17(8), 1332.
[17] Sun, T., Yang, J., Li, J., Chen, J., Liu, M., Fan, L., & Wang, X. (2024). Enhancing auto insurance risk evaluation with transformer and SHAP. IEEE Access.
[18] Helton, M., Zerafa, S., Vishwanath, K., & Mycek, M. A. (2022). Efficient computation of the steady-state and time-domain solutions of the photon diffusion equation in layered turbid media. Scientific Reports, 12(1), 18979.
[19] Ma, Z., Chen, X., Sun, T., Wang, X., Wu, Y. C., & Zhou, M. (2024). Blockchain-based zero-trust supply chain security integrated with deep reinforcement learning for inventory optimization. Future Internet, 16(5), 163.
[20] Mai, N. T., Cao, W., & Liu, W. (2025). Interpretable knowledge tracing via transformer-Bayesian hybrid networks: Learning temporal dependencies and causal structures in educational data. Applied Sciences, 15(17), 9605.

ISSN: 3079-9325

[21] Qiu, L. (2025). Reinforcement Learning Approaches for Intelligent Control of Smart Building Energy Systems with Real-Time Adaptation to Occupant Behavior and Weather Conditions. Journal of Computing and Electronic Information Management, 18(2), 32-37.
[22] Zhang, C., & Wang, X. (2020). Quasi-Monte Carlo-based conditional pathwise method for option Greeks. Quantitative Finance, 20(1), 49-67.
[23] Lin, J. A. (2025). Scalable Gaussian Processes: Advances in Iterative Methods and Pathwise Conditioning. arXiv preprint arXiv:2507.06839.
[24] Zhang, H. (2025). Physics-Informed Neural Networks for High-Fidelity Electromagnetic Field Approximation in VLSI and RF EDA Applications. Journal of Computing and Electronic Information Management, 18(2), 38-46.
[25] Qiu, L. (2025). Multi-Agent Reinforcement Learning for Coordinated Smart Grid and Building Energy Management Across Urban Communities. Computer Life, 13(3), 8-15.
[26] Dossatayev, A., Manapova, A., & Omarov, B. (2023). Deep residual convolutional long short-term memory network for option price prediction problem. International Journal of Advanced Computer Science and Applications, 14(9).
[27] Pohjonen, T. (2022). Pricing and Hedging Options with Neural Networks: A Case of FTSE 100 Index Options.
[28] Li, J., Fan, L., Wang, X., Sun, T., & Zhou, M. (2024). Product demand prediction with spatial graph neural networks. Applied Sciences, 14(16), 6989.
[29] Amir, A. K. (2023). Analyzing Numerical Methods for Solving Telegraph Partial Differential Equations: A Comprehensive Literature Review on Difference Scheme Approaches. Available at SSRN 4641627.
[30] Qiu, L. (2025). Machine Learning Approaches to Minimize Carbon Emissions through Optimized Road Traffic Flow and Routing. Frontiers in Environmental Science and Sustainability, 2(1), 30-41.
[31] Boudardara, F., Boussif, A., Meyer, P. J., & Ghazel, M. (2024). A review of abstraction methods toward verifying neural networks. ACM Transactions on Embedded Computing Systems, 23(4), 1-19.
[32] Armeniakos, G., Zervakis, G., Soudris, D., & Henkel, J. (2022). Hardware approximate techniques for deep neural network accelerators: A survey. ACM Computing Surveys, 55(4), 1-36.
[33] Gomes, P. D. (2024). Mathematics of differential machine learning in derivative pricing and hedging. arXiv preprint arXiv:2405.01233.
[34] Voskou, A. (2024). Deep Transformer Neural Networks with Stochastic Competition (Doctoral dissertation).
[35] Ter-Avanesov, B., & Beigi, H. (2024). MLP, XGBoost, KAN, TDNN, and LSTM-GRU Hybrid RNN with Attention for SPX and NDX European Call Option Pricing. arXiv preprint arXiv:2409.06724.
[36] Neelan, A. G., Bosco, F. S. D., Jarugumalli, N. S., & Vedarethinam, S. B. (2025). Physics-Informed Neural Networks: Bridging the Divide Between Conservative and Non-Conservative Equations. arXiv preprint arXiv:2506.22413.
[37] Ge, Y., Wang, Y., Liu, J., & Wang, J. (2025). GAN-Enhanced Implied Volatility Surface Reconstruction for Option Pricing Error Mitigation. IEEE Access.
[38] Usama, M., Ahmad, B., Song, E., Hossain, M. S., Alrashoud, M., & Muhammad, G. (2020). Attention-based sentiment analysis using convolutional and recurrent neural network. Future Generation Computer Systems, 113, 571-578.-54649.
[39] Mienye, I. D., Swart, T. G., & Obaido, G. (2024). Recurrent neural networks: A comprehensive review of architectures, variants, and applications. Information, 15(9), 517.
[40] Borup, D., & Jakobsen, J. S. (2019). Capturing volatility persistence: A dynamically complete realized EGARCH-MIDAS model. Quantitative Finance, 19(11), 1839-1855.