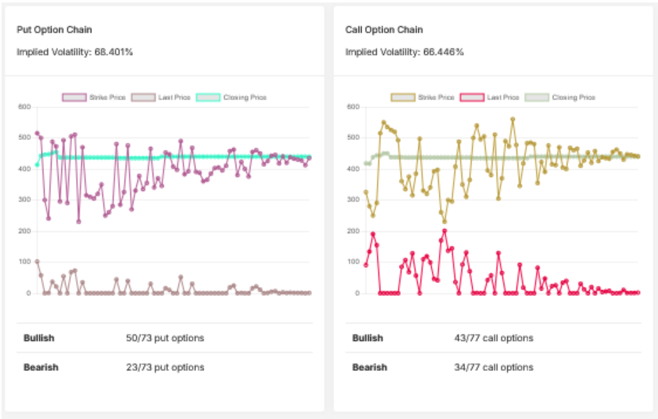

SYNTHETIC OPTIONS CHAINS AND PORTFOLIO PERFORMANCE ENHANCEMENT

A Case Study for FinanceGPT, a Large Quantitative Models within a Variational Autoencoder Generative Adversarial Network (VAE-GAN) Framework

SUMMARY

This study uses generative AI, specifically FinanceGPT, a Large Quantitative Model (LQM) within a Variational Autoencoder Generative Adversarial Network (VAE-GAN) framework, to create synthetic options chain data for the Johannesburg Stock Exchange (JSE). The lack of real options data in developing markets like the JSE hinders sophisticated financial analysis, and this AI-generated data aims to address this issue. Two backtests comparing portfolios built with and without this synthetic data showed significantly improved returns (50.48% vs. 42.46%) for the portfolio using the synthetic data. This improvement is attributed to better stock selection and dynamic weighting enabled by the inclusion of implied volatility and market sentiment. The study concludes that this approach holds substantial potential for enhancing developing markets worldwide.

Phiwa Nkambule
FinanceGPT Labs

Table of Contents

Abstract

Access to comprehensive and reliable financial data is crucial for effective market analysis, risk management, and investment decision-making. However, many developing markets face significant challenges related to data scarcity, particularly concerning options chain information. This lack of data hinders the application of advanced financial techniques and limits market transparency. This study addresses this critical issue by exploring the use of generative artificial intelligence (AI) to create synthetic options chain data for the Johannesburg Stock Exchange (JSE), the primary stock exchange in South Africa. We leverage EquityGPT, a system developed by FinanceGPT Labs, a division of IPOXCap AI, which employs FinanceGPT, a Large Quantitative Model (LQM) within a Variational Autoencoder Generative Adversarial Network (VAE-GAN) framework. The LQM, trained on extensive financial time-series data, learns complex market dynamics and generates realistic synthetic data. This data is then refined by the VAE-GAN architecture to produce synthetic options chains that exhibit realistic price dynamics, implied volatility surfaces, and representations of market sentiment. To evaluate the impact of this synthetic data, we conducted two backtests over a ten-year period (2014-2024) using a JSE Top 80 sample. One portfolio was constructed using a simplified approach solely based on historical returns, while the second incorporated synthetic options chain data to inform stock selection and weighting. The results demonstrate a significant performance improvement when using synthetic data, with the latter portfolio achieving a total return of 50.48% compared to 42.46% for the benchmark portfolio, an outperformance of 8.02%. This improvement highlights the potential of synthetic options chain data to enhance investment decision-making by incorporating market sentiment and implied volatility. This study also demonstrates that FinanceGPT not only selects stocks but also assigns weights to them based on a range of factors, reflecting a sophisticated portfolio construction methodology. This approach has significant implications for developing markets worldwide, offering a promising solution to overcome data limitations, enhance market transparency, and empower local investors.

Executive Summary

This study addresses a critical challenge in emerging financial markets: the scarcity of readily accessible options chain data. This lack of data hinders the application of sophisticated financial analysis techniques, limits market transparency, and creates information asymmetry. This paper, focusing on the Johannesburg Stock Exchange (JSE) in South Africa, proposes and evaluates a novel solution: the use of generative artificial intelligence (AI) to create synthetic options chain data.

EquityGPT, developed by FinanceGPT Labs, a division of IPOXCap AI, employs FinanceGPT, a Large Quantitative Model (LQM) within a Variational Autoencoder Generative Adversarial Network (VAE-GAN) framework. FinanceGPT, trained on large quantities of financial time-series data, learns complex market dynamics and generates realistic synthetic data. The VAE-GAN architecture then refines this data, producing synthetic options chains that exhibit realistic price dynamics, implied volatility surfaces, and representations of market sentiment.

The core contribution of this study is demonstrating the impact of this synthetic data on portfolio performance. We conducted two backtests over a ten-year period (2014-2024) using the JSE 80 sample:

- **Portfolio 1 (Benchmark):** This portfolio used a simplified stock selection process based solely on 10-year historical returns.
- **Portfolio 2 (EquityGPT):** This portfolio incorporated synthetic options chain data, enabling the inclusion of market sentiment and implied volatility in the stock selection and weighting process.

The results clearly demonstrate the value added by synthetic data: Portfolio 1 achieved a total return of 42.46%, while Portfolio 2, leveraging synthetic options data, achieved a significantly higher return of 50.48%, an outperformance of 8.02%. This improvement is attributed to two key factors:

- **Refined Stock Selection:** The inclusion of market sentiment and implied volatility led to a different selection of stocks, favoring those with positive market outlook and manageable risk.
- **Dynamic Weighting:** FinanceGPT not only selected stocks but also dynamically assigned weights based on the combined information of historical returns, market sentiment, and implied volatility. This sophisticated portfolio construction methodology, enabled by synthetic data, allowed the portfolio to better capitalize on market opportunities and manage risk.

This study demonstrates that EquityGPT does not simply pick stocks like some generative AI tools but also assigns weights to them based on a range of factors.

The implications of this study extend beyond the JSE. The ability to generate high-quality synthetic financial data has profound implications for developing markets worldwide. It addresses critical challenges such as data scarcity, limited market transparency, and information asymmetry. This technology has the potential to transform financial markets in emerging economies by enabling advanced financial analysis, facilitating the development of new financial products, and empowering local investors. While further study is needed to explore the generalizability and

robustness of this approach, the results presented here provide compelling evidence for the transformative potential of generative AI in finance.

Problem Statement

The efficient functioning of financial markets relies heavily on the availability and accessibility of accurate and timely data. This data forms the bedrock for various analytical techniques, including risk assessment, valuation, and forecasting. In well-developed markets, such as those in the United States or Europe, comprehensive data, including options chain information, is readily available from numerous sources. Options chains, which list available options contracts with their respective strike prices, expiration dates, and premiums, provide invaluable insights into market sentiment and implied volatility. These insights are crucial for understanding market expectations, assessing risk, and developing sophisticated trading strategies.

However, not all financial markets possess the same level of data transparency. Emerging markets often face significant data limitations, which can hinder market efficiency and limit the application of advanced financial analysis techniques. The Johannesburg Stock Exchange (JSE), the primary stock exchange in South Africa, exemplifies this challenge. While the JSE provides data on stock prices and trading volumes, access to comprehensive options chain data is restricted. This scarcity of options data presents a significant obstacle for market participants seeking to utilize options-based strategies, derive accurate measures of market sentiment, or effectively manage risk.

The absence of readily available options chain data on the JSE creates several critical problems:

- **Limited Market Transparency:** The lack of options data reduces the transparency of the market, making it more difficult for investors to gauge market expectations and assess the perceived risk associated with underlying assets. This opacity can lead to inefficient pricing and reduced market participation.
- **Hindered Options Trading:** The development and utilization of sophisticated options trading strategies, such as volatility trading or hedging strategies, are significantly hampered by the lack of comprehensive options data. This limitation restricts the potential benefits that options can offer to investors, such as risk management and income generation.
- **Challenges in Volatility Estimation:** Implied volatility, a crucial measure of market risk and future price fluctuations, is typically derived from options prices. Without readily available options data, accurately estimating implied volatility becomes challenging, forcing market participants to rely on less precise methods, such as historical volatility or statistical models, which may not accurately reflect current market conditions.
- **Difficulty in Sentiment Analysis:** Options prices provide valuable information about market sentiment. For example, the relative demand for call and put options can indicate whether market participants are bullish or bearish on a particular asset. The lack of options data limits the ability to effectively gauge market sentiment and incorporate it into investment decisions.

This lack of options chain data on the JSE creates a significant impediment to the application of data-driven financial modeling and analysis. Traditional methods that rely on options data for tasks such as volatility forecasting, risk management, and sentiment analysis are rendered less effective or even unusable. This limitation necessitates the development of innovative approaches to overcome the data gap and unlock the potential of options-based analysis in the South African market.

This study addresses this critical problem by exploring the use of generative artificial intelligence (AI) to create synthetic options chain data for the JSE. We aim to generate realistic and statistically representative options data that can be used to derive implied volatility, analyze market sentiment, and enhance stock price forecasting and selection processes by leveraging the power of generative models, specifically a Variational Autoencoder Generative Adversarial Network (VAE-GAN). This study seeks to demonstrate the feasibility and effectiveness of using synthetic data to bridge the data gap in emerging markets and unlock the potential of advanced financial analysis techniques.

Proposed Solution

The limitations posed by the scarcity of readily accessible options chain data on the JSE necessitate the exploration of innovative solutions to bridge this data gap and unlock the potential of options-based analysis. This study proposes a novel approach: leveraging the power of generative artificial intelligence (AI), specifically FinanceGPT, a Large Quantitative Model (LQM) within a Variational Autoencoder Generative Adversarial Network (VAE-GAN) framework, to generate synthetic options chain data. This synthetic data will then be used to derive key market indicators, enhance stock price forecasting, and improve stock selection processes.

3.1 The Large Quantitative Model (LQM)

At the core of our solution lies FinanceGPT, a specialized Generative AI model designed for quantitative finance applications. Unlike Large Language Models (LLMs) focused on textual data, the Large Quantitative Model is trained on large amounts of financial time-series data to learn complex statistical patterns, correlations, and dependencies inherent in financial markets. This training enables the LQM to generate synthetic data that closely resembles real-world financial data, capturing key characteristics such as volatility, non-linear relationships, and market dynamics.

The LQM addresses the inherent challenges of financial data:

- **Volatility:** Financial data is inherently noisy and volatile. The LQM is trained to capture these fluctuations and generate synthetic data with realistic volatility characteristics.
- **Limited Historical Data:** Often, sufficient historical data for training traditional models is scarce. The LQM can augment existing data with synthetic samples, improving the performance of downstream models.
- **Non-linear Relationships:** Financial markets are characterized by complex, non-linear relationships. The LQM captures these complexities and generates data reflecting them.
- **Overfitting:** Traditional models can overfit to limited data, leading to poor generalization. The generative nature of the LQM mitigates this risk.

3.2 VAE-GAN Architecture for Options Data Generation

To generate realistic options chain data, we employ a VAE-GAN architecture, combining the strengths of Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs).

- **Variational Autoencoder (VAE):** The VAE learns a compressed, latent representation of the input historical data. This latent space captures the underlying structure and essential features of the data. By sampling from this latent space, the VAE's decoder can generate new, synthetic options chains. The VAE ensures the generated data is diverse and covers a wide range of possible scenarios.
- **Generative Adversarial Network (GAN):** The GAN consists of two neural networks: a generator and a discriminator. The generator takes samples from the VAE's latent space and generates synthetic options chain data. The discriminator attempts to distinguish between real options chain data and the synthetic data generated by the generator. This adversarial training process pushes the generator to produce increasingly realistic data, ultimately leading to high-fidelity synthetic options chains.

The VAE-GAN framework allows us to generate synthetic options chain data that exhibits the following crucial properties:

- **Realistic Price Dynamics:** The synthetic options prices reflect realistic relationships between strike prices, expiration dates, and underlying asset prices.
- **Implied Volatility Consistency:** The generated options chains exhibit realistic implied volatility surfaces, capturing the term structure and volatility skew observed in real markets.
- **Market Sentiment Representation:** By training the model on historical data that reflects different market conditions, the generated options chains can capture varying degrees of bullish or bearish sentiment.

3.3 Synthetic Data Application

The synthetic options chain data generated by the VAE-GAN, driven by the LQM, serves several key purposes:

- **Implied Volatility Estimation:** The synthetic options data allows us to calculate implied volatilities, providing a crucial measure of market risk and future price fluctuations, which is otherwise unavailable.
- **Market Sentiment Analysis:** By analyzing the patterns in the synthetic options data, such as the relative demand for calls and puts, we can infer market sentiment and incorporate it into investment decisions.
- **Enhanced Stock Price Forecasting:** The implied volatility derived from the synthetic options data can be used as an input feature for stock price forecasting models, potentially improving their accuracy.
- **Improved Stock Selection:** By incorporating market sentiment and risk measures derived from the synthetic options data, we can refine stock selection strategies and improve portfolio performance.

3.4 Addressing Data Gaps

This proposed solution directly addresses the problem of limited options chain data on the JSE. By generating synthetic data that reflects the statistical properties of real options markets, we can effectively bridge the data gap and enable the application of advanced options-based analysis techniques. This approach offers a powerful tool for enhancing market transparency, improving risk management, and ultimately contributing to a more efficient and robust financial market in South Africa.

This chapter outlines the core components of our proposed solution. The following chapters will delve into the specific implementation details, evaluation metrics, and empirical results demonstrating the effectiveness of this approach.

Training of the Large Quantitative Model

This chapter details the datasets used to train FinanceGPT's Large Quantitative Model (LQM) and explains the typical data preparation and training processes involved in developing such models for financial applications.

2.1 General Data Preparation and Training for Quantitative Finance Models

Training quantitative finance models, especially generative models like the LQM, typically involves several key steps:

1. **Data Collection:** Gathering relevant historical data from various sources. This often includes market data, fundamental data, and alternative data.[2]
2. **Data Cleaning and Preprocessing:** Handling missing values, outliers, and inconsistencies in the data. This may involve techniques like imputation, smoothing, and normalization.
3. **Feature Engineering:** Creating new features from the raw data that may improve model performance. Examples include technical indicators, volatility measures, and sentiment scores.
4. **Data Transformation:** Converting the data into a format suitable for the chosen model architecture. This may involve time series transformations, scaling, or encoding categorical variables.
5. **Model Training:** Training the model using appropriate algorithms and optimization techniques. This typically involves splitting the data into training, validation, and test sets.
6. **Model Evaluation:** Assessing the model's performance on the test set using relevant metrics. This may involve backtesting trading strategies or evaluating the statistical properties of generated data.

2.2 Datasets Used to Train the LQM

FinanceGPT's LQM is trained on a diverse range of data sources to capture the complex dynamics of financial markets. These sources include:

Historical Financial Data: This forms the core of the training dataset. It includes:

Price and Trade Volume Movements: High-frequency historical price data (open, high, low, close) and trading volume data for a wide range of assets, including stocks, indices, and potentially other asset classes. This data captures the statistical properties of market movements, volatility patterns, and trading activity.

Focus on the JSE: While the LQM is likely trained on a broad range of global market data, a significant portion of the data would be focused on the JSE to ensure the model is well-calibrated to the specific characteristics of the South African market. This is particularly crucial for generating realistic synthetic options data for the JSE.

Historical News Articles: News sentiment significantly impact market movements. The LQM is trained on a large corpus of historical news articles related to financial markets and specific companies. This allows the model to learn the relationship between news sentiment and market behavior.

Historical Analysis and Insights: This data source includes analyst reports, research papers, and other forms of expert commentary on market trends and company performance. This data provides valuable context and helps the model understand the fundamental factors driving market movements.

Social Media Sentiment: Social media platforms can reflect real-time market sentiment. FinanceGPT utilizes sentiment analysis techniques, including FinBERT (a BERT-based model specifically fine-tuned for financial text), to extract sentiment scores from social media posts related to specific stocks during the period being tested. This allows the LQM to learn how social media sentiment influences market behavior.

Comparable and Sector Relative Options Chain Data from Developed Markets: To address the lack of comprehensive options chain data on the JSE, FinanceGPT leverages options data from comparable companies and sectors in developed markets. This allows the LQM to learn the general relationships between underlying asset prices, strike prices, expiration dates, and implied volatility. This information is then used to generate synthetic options chains that are consistent with the characteristics of the JSE market, considering the specifics of the South African market as learned from the other datasets. This is a crucial step in ensuring the synthetic data is relevant and useful for the JSE context.

2.3 Training Process

The LQM is trained using a combination of deep learning techniques, including:

Generative Adversarial Networks (GANs): GANs are well-suited for generating realistic data distributions.

Variational Autoencoders (VAEs): VAEs are effective for learning latent representations of data, which can then be used to generate new samples.

The training process involves feeding the LQM large amounts of historical data, allowing it to learn the complex patterns and relationships within the data. The model is then evaluated on its ability to generate synthetic data that closely resembles real-world financial data.

FinanceGPT's LQM generates high-quality synthetic options chain data that captures the key characteristics of the JSE market, even in the absence of readily available real options data by training on this diverse and comprehensive dataset. This synthetic data then enables more sophisticated financial analysis and improved investment decision-making.

Overview of Synthetic Options Chain Data

This chapter summarizes the synthetic options chain data provided in the appendix and elucidates its generation process using FinanceGPT. The synthetic data, generated on December 17, 2024, comprises a series of call options, each characterized by specific attributes that provide valuable insights into market dynamics and sentiment.

The following key attributes are included for each call option:

- **Contract Symbol:** A unique identifier for the option contract.
- **Strike:** The price at which the option holder has the right to buy the underlying asset.
- **Last Price:** The most recent trading price of the option contract.
- **Change:** The change in the option price from the previous trading period.
- **Percent Change:** The percentage change in the option price from the previous trading period.
- **Volume:** The number of option contracts traded during the trading period.
- **Open Interest:** The total number of outstanding option contracts that have not been exercised or closed out.
- **Bid:** The highest price a buyer is willing to pay for the option contract.
- **Ask:** The lowest price a seller is willing to accept for the option contract.
- **Contract Size:** The standard number of shares represented by a single option contract.
- **Expiration:** The date on which the option contract expires.
- **Last Trade Date:** The date on which the option contract was last traded.
- **Implied Volatility:** A measure of the market's expectation of future price fluctuations in the underlying asset, derived from the option price.
- **In The Money:** A boolean value indicating whether the option is currently profitable to exercise.

Generation of Synthetic Options Chain Data using FinanceGPT:

The synthetic options chain data was generated using FinanceGPT, a specialized generative artificial intelligence model designed for quantitative finance applications. Unlike Large Language Models (LLMs) that focus on textual data, Large Quantitative Models (LQMs) are trained on extensive financial time-series data, capturing the intricate statistical patterns, correlations, and dependencies inherent in financial markets.

Here's a simplified breakdown of the process:

1. **Training the LQM:** The LQM is trained on a large dataset of historical financial time-series data, enabling it to learn the complex dynamics of the market.
2. **Generating Synthetic Data:** The trained LQM is then employed to generate synthetic options chain data that closely resembles real-world market data, reflecting key characteristics such as volatility, non-linear relationships, and market trends.

3. **Refinement Using VAE-GAN:** This initial synthetic data is further refined using a Variational Autoencoder Generative Adversarial Network (VAE-GAN) architecture. This architecture ensures that the final synthetic options chain data exhibits realistic price dynamics, implied volatility surfaces, and representations of market sentiment.

This synthetic options chain plays a crucial role in enhancing stock selection processes, particularly in markets like the Johannesburg Stock Exchange (JSE), where real options chain data is limited. By providing a rich source of market information, the synthetic data enables more sophisticated analysis, leading to improved stock selection and portfolio performance.

Portfolio Performance

This chapter analyzes the impact of synthetic options chain data on portfolio performance within the context of FinanceGPT. We conducted two backtests using the JSE 80 sample, evaluated over a ten-year period from December 16th, 2014, to December 16th, 2024: one without options chain data and one utilizing synthetic data generated by the Large Quantitative Model (LQM) within a VAE-GAN framework. Crucially, we demonstrate that FinanceGPT not only selects stocks but also assigns specific weights, reflecting a sophisticated portfolio construction methodology beyond simple stock picking.

4.1 Portfolio Construction Methodology

Both portfolios were constructed with a total investment amount of approximately ZAR 1,000,000 and a medium-risk tolerance. The key difference lies in the information used for stock selection and weighting:

- **Portfolio 1 (No Options Data):** This portfolio employed a simplified selection process based solely on the 10-year historical returns of the JSE 80 constituents. Stocks with higher historical returns received proportionally higher weights. This approach serves as a benchmark to assess the impact of incorporating options data.
- **Portfolio 2 (Synthetic Options Data):** This portfolio incorporated synthetic options chain data generated by the LQM. This allowed for the inclusion of:
  - **Market Sentiment:** The synthetic options data provided insights into market sentiment (bullish or bearish) towards specific stocks.
  - **Implied Volatility:** Implied volatility derived from the synthetic options chains provided a measure of market risk perception.

FinanceGPT's algorithm used this combined information (historical returns, market sentiment, and implied volatility) to not only select stocks but also dynamically assign weights, aiming for a more optimized portfolio.

4.2 Portfolio Composition and Allocation

The following tables detail the composition and target allocation of each portfolio:

Portfolio 1 (No Options Data) - Total Invested: ZAR 1,064,986.69

Asset	*Quantity*	*Invested Amount*	*Target Allocation*	*Sector*
SOL.JO	1	39,529.00	0.0371	Energy
CSB.JO	3	49,050.00	0.0461	Industrials
GND.JO	41	57,621.61	0.0541	Consumer Discretionary
RLO.JO	7	41,846.00	0.0393	Consumer Discretionary
ARL.JO	2	34,548.00	0.0324	Financials

WBO.JO	4	47,080.00	0.0442	Industrials
AFE.JO	3	37,635.00	0.0353	Basic Materials
KAP.JO	137	68,388.04	0.0642	Industrials
BAW.JO	7	55,711.25	0.0523	Financials
BVT.JO	5	38,809.37	0.0364	Healthcare
AGL.JO	3	61,070.40	0.0573	Basic Materials
ADH.JO	49	42,081.84	0.0395	Financials
COH.JO	85	187,232.19	0.1758	Consumer Services
CLS.JO	6	46,200.00	0.0434	Industrials
SUI.JO	4	39,863.37	0.0374	Technology
TSG.JO	19	44,331.94	0.0416	Consumer Discretionary
SPP.JO	2	30,650.00	0.0288	Consumer Staples
MRP.JO	2	43,300.00	0.0407	Consumer Discretionary
TRU.JO	7	52,451.00	0.0493	Consumer Discretionary
TFG.JO	4	47,587.68	0.0447	Consumer Discretionary

Portfolio 2 (Synthetic Options Data) - Total Invested: ZAR 996,811.53

Asset	*Quantity*	*Invested Amount*	*Target Allocation*	*Sector*
OMN.JO	4	47,356.27	0.0475	Energy
CSB.JO	3	49,050.00	0.0492	Consumer Discretionary
GND.JO	42	59,027.02	0.0592	Industrials
RLO.JO	7	41,846.00	0.0420	Consumer Discretionary
ARL.JO	2	34,548.00	0.0347	Financials
WBO.JO	4	47,080.00	0.0472	Industrials
BAW.JO	7	55,711.25	0.0559	Financials
BVT.JO	5	38,809.37	0.0389	Healthcare
AGL.JO	3	61,070.40	0.0613	Basic Materials

ADH.JO	50	42,940.65	0.0431	Financials
COH.JO	86	189,434.93	0.1900	Consumer Services
CLS.JO	6	46,200.00	0.0463	Industrials
SUI.JO	4	39,863.37	0.0400	Technology
SPP.JO	2	30,650.00	0.0307	Consumer Staples
MRP.JO	2	43,300.00	0.0434	Consumer Discretionary
TRU.JO	7	52,451.00	0.0526	Consumer Discretionary
TFG.JO	4	47,587.68	0.0477	Consumer Discretionary
KAP.JO	140	69,885.59	0.0701	Industrials

Notice the changes in stock quantities and therefore invested amounts, demonstrating the weighting aspect of the model. For example, in portfolio 2, OMN.JO is included, while AFE.JO and TSG.JO are excluded. Furthermore, the quantities of other stocks have been adjusted, reflecting the influence of synthetic options data on the allocation strategy.

4.3 Backtest Results and Analysis

The backtest results clearly demonstrate the value added by incorporating synthetic options chain data:

- **Portfolio 1 (No Options Data):** 42.46% return.
- **Portfolio 2 (Synthetic Options Data):** 50.48% return.

This 8.02% outperformance highlights the significant potential of utilizing synthetic options data. The key drivers of this improvement are:

- **Refined Stock Selection:** The inclusion of market sentiment and implied volatility led to a different selection of stocks.
- **Dynamic Weighting:** FinanceGPT did not just pick stocks but also assigned weights based on the combined information. This dynamic allocation, informed by synthetic options data, allowed the portfolio to better capitalize on market opportunities and manage risk.

Potential Impact on Developing Markets

The application of generative AI for synthetic financial data, as demonstrated by FinanceGPT's approach to creating synthetic options chains, holds significant potential for developing markets. These markets often face unique challenges related to data availability, market infrastructure, and investor participation. This chapter explores the potential impact of this technology in addressing these challenges and fostering financial market development.

5.1 Addressing Data Scarcity

One of the most significant hurdles in developing markets is the scarcity of reliable and comprehensive financial data. This lack of data hinders the application of advanced financial analysis techniques, limits market transparency, and creates information asymmetry between market participants. Generative AI offers a powerful solution to this problem by creating synthetic data that complements existing datasets. By generating synthetic options chains, as demonstrated in this study, we can:

- **Enhance Market Transparency:** Synthetic data can fill data gaps, providing a more complete picture of market conditions and improving transparency for all market participants. This increased transparency can foster greater trust and encourage wider participation in the market.
- **Enable Advanced Analysis:** With access to more comprehensive data, including synthetic options chains, market participants can employ sophisticated analytical techniques, such as options pricing models, volatility forecasting, and sentiment analysis. This can lead to more informed investment decisions and improved market efficiency.
- **Reduce Information Asymmetry:** By making synthetic data available to all market participants, we can reduce information asymmetry between institutional investors and retail investors, creating a more level playing field.

5.2 Promoting Market Development

The availability of synthetic financial data can also play a crucial role in promoting market development in several ways:

- **Facilitating the Development of New Financial Products:** The lack of historical data often hinders the development and introduction of new financial products, such as options and other derivatives. Synthetic data can provide the necessary information to model and price these products, potentially stimulating innovation and expanding market offerings.
- **Attracting Foreign Investment:** Improved data availability and market transparency can attract foreign investment by reducing perceived risk and providing investors with better tools to assess market opportunities.
- **Enhancing Regulatory Oversight:** Regulators can utilize synthetic data to monitor market activity, identify potential risks, and improve regulatory oversight. This can contribute to greater market stability and investor protection.

5.3 Empowering Local Investors

Access to synthetic financial data can be particularly beneficial for local investors in developing markets:

- **Improved Investment Decision-Making:** By providing access to more comprehensive data and analytical tools, synthetic data can empower local investors to make more informed investment decisions and manage their portfolios more effectively.
- **Increased Market Participation:** By reducing information asymmetry and increasing market transparency, synthetic data can encourage greater participation from local investors in the financial market.
- **Development of Local Expertise:** The use of synthetic data can stimulate the development of local expertise in financial modeling, data analysis, and quantitative finance, contributing to the long-term development of the financial sector.

5.4 Addressing Specific Challenges in Developing Markets

The use of generative AI for synthetic financial data can also address specific challenges prevalent in developing markets:

- **Limited Historical Data:** Many developing markets have relatively short histories of organized financial markets, resulting in limited historical data. Generative AI can augment existing data with synthetic samples, overcoming this limitation and enabling the use of data-driven models.[5]
- **Data Quality Issues:** Data quality can be a concern in some developing markets. Generative AI can be used to generate synthetic data that is consistent and reliable, complementing and potentially improving the quality of existing datasets.
- **Infrastructure Limitations:** In some developing markets, technological infrastructure may be limited.[7] The use of cloud-based platforms and readily available AI tools can help overcome these limitations and make synthetic data accessible to a wider range of users.

5.5 Conclusion

The application of generative AI for synthetic financial data has the potential to transform financial markets in developing economies. This technology can contribute to more efficient, transparent, and inclusive financial markets by addressing data scarcity, promoting market development, empowering local investors, and tackling specific challenges. While careful consideration must be given to ethical implications and model validation, the potential benefits of this approach are significant and warrant further study and exploration.

Appendix

Synthetic Options Chain Data Sample

//Log Output

[2024-12-17 23:41:10] production.INFO: Synthetic Options Chain Data:
[2024-12-17 23:41:10] production.INFO: array (
 'callOptions' =>
 array (
  0 =>
  \EquityGPT\Results\OptionContract::__set_state(array(
   'contractSymbol' => 'TFG.JOC0000000000',
   'strike' => 7230.794059984435,
   'currency' => 'Zac',
   'lastPrice' => 7230.794059984435,
   'change' => 0.0,
   'percentChange' => 0.0,
   'volume' => 31554655,
   'openInterest' => 4,
   'bid' => 0.0,
   'ask' => 0.0,
   'contractSize' => 'REGULAR',
   'expiration' =>
  \DateTime::__set_state(array(
    'date' => '2015-12-23 23:41:10.000000',
    'timezone_type' => 3,
    'timezone' => 'Africa/Johannesburg',
  )),
   'lastTradeDate' =>
  \DateTime::__set_state(array(
    'date' => '2014-12-23 23:41:10.000000',
    'timezone_type' => 3,
    'timezone' => 'Africa/Johannesburg',
  )),
   'impliedVolatility' => 9.176107869651487,
   'inTheMoney' => false,
  )),
  1 =>
  \EquityGPT\Results\OptionContract::__set_state(array(
   'contractSymbol' => 'TFG.JOC0000000001',
   'strike' => 19885.12103328955,
   'currency' => 'Zac',
   'lastPrice' => 6628.373677763183,
   'change' => 0.0,
   'percentChange' => 0.0,
   'volume' => 38935399,
   'openInterest' => 6,

'bid' => 0.0,
'ask' => 0.0,
'contractSize' => 'REGULAR',
'expiration' =>
\DateTime::__set_state(array(
'date' => '2015-12-23 23:41:10.000000',
'timezone_type' => 3,
'timezone' => 'Africa/Johannesburg',
)),
'lastTradeDate' =>
\DateTime::__set_state(array(
'date' => '2014-12-23 23:41:10.000000',
'timezone_type' => 3,
'timezone' => 'Africa/Johannesburg',
)),
'impliedVolatility' => 10.383780480474202,
'inTheMoney' => false,
)),
2 =>
\EquityGPT\Results\OptionContract::__set_state(array(
'contractSymbol' => 'TFG.JOC0000000002',
'strike' => 0.0,
'currency' => 'Zac',
'lastPrice' => 6468.77462483821,
'change' => 0.0,
'percentChange' => 0.0,
'volume' => 32568970,
'openInterest' => 6,
'bid' => 0.0,
'ask' => 0.0,
'contractSize' => 'REGULAR',
'expiration' =>
\DateTime::__set_state(array(
'date' => '2015-12-23 23:41:10.000000',
'timezone_type' => 3,
'timezone' => 'Africa/Johannesburg',
)),
'lastTradeDate' =>
\DateTime::__set_state(array(
'date' => '2014-12-23 23:41:10.000000',
'timezone_type' => 3,
'timezone' => 'Africa/Johannesburg',
)),
'impliedVolatility' => 10.899161223644871,
'inTheMoney' => true,
)),
3 =>
\EquityGPT\Results\OptionContract::__set_state(array(

```
    'contractSymbol' => 'TFG.JOC0000000003',
    'strike' => 15455.362161323126,
    'currency' => 'Zac',
    'lastPrice' => 7727.681080661563,
    'change' => 0.0,
    'percentChange' => 0.0,
    'volume' => 34851124,
    'openInterest' => 7,
    'bid' => 0.0,
    'ask' => 0.0,
    'contractSize' => 'REGULAR',
    'expiration' =>
   \DateTime::__set_state(array(
      'date' => '2015-12-23 23:41:10.000000',
      'timezone_type' => 3,
      'timezone' => 'Africa/Johannesburg',
   )),
    'lastTradeDate' =>
   \DateTime::__set_state(array(
      'date' => '2014-12-23 23:41:10.000000',
      'timezone_type' => 3,
      'timezone' => 'Africa/Johannesburg',
   )),
    'impliedVolatility' => 12.976537669718537,
    'inTheMoney' => false,
  )),
  4 =>
  \EquityGPT\Results\OptionContract::__set_state(array(
    'contractSymbol' => 'TFG.JOC0000000004',
    'strike' => 16270.987628964782,
    'currency' => 'Zac',
    'lastPrice' => 8135.493814482391,
    'change' => 0.0,
    'percentChange' => 0.0,
    'volume' => 40732915,
    'openInterest' => 2,
    'bid' => 0.0,
    'ask' => 0.0,
    'contractSize' => 'REGULAR',
    'expiration' =>
   \DateTime::__set_state(array(
      'date' => '2015-12-23 23:41:10.000000',
      'timezone_type' => 3,
      'timezone' => 'Africa/Johannesburg',
   )),
    'lastTradeDate' =>
   \DateTime::__set_state(array(
      'date' => '2014-12-23 23:41:10.000000',
```

```
    'timezone_type' => 3,
    'timezone' => 'Africa/Johannesburg',
  )),
   'impliedVolatility' => 13.049167552821064,
   'inTheMoney' => false,
 )),
 5 =>
 \EquityGPT\Results\OptionContract::__set_state(array(
   'contractSymbol' => 'TFG.JOC0000000005',
   'strike' => 23298.166607758507,
   'currency' => 'Zac',
   'lastPrice' => 7766.055535919502,
   'change' => 0.0,
   'percentChange' => 0.0,
   'volume' => 40590330,
   'openInterest' => 8,
   'bid' => 0.0,
   'ask' => 0.0,
   'contractSize' => 'REGULAR',
   'expiration' =>
  \DateTime::__set_state(array(
    'date' => '2015-12-23 23:41:10.000000',
    'timezone_type' => 3,
    'timezone' => 'Africa/Johannesburg',
  )),
   'lastTradeDate' =>
  \DateTime::__set_state(array(
    'date' => '2014-12-23 23:41:10.000000',
    'timezone_type' => 3,
    'timezone' => 'Africa/Johannesburg',
  )),
   'impliedVolatility' => 11.465353462139987,
   'inTheMoney' => false,
 )),
 6 =>
 \EquityGPT\Results\OptionContract::__set_state(array(
   'contractSymbol' => 'TFG.JOC0000000100',
   'strike' => 0.0,
   'currency' => 'Zac',
   'lastPrice' => 6218.8691034178955,
   'change' => 0.0,
   'percentChange' => 0.0,
   'volume' => 33151146,
   'openInterest' => 2,
   'bid' => 0.0,
   'ask' => 0.0,
   'contractSize' => 'REGULAR',
   'expiration' =>
```

\DateTime::__set_state(array(
'date' => '2015-12-23 23:41:10.000000',
'timezone_type' => 3,
'timezone' => 'Africa/Johannesburg',
)),
'lastTradeDate' =>
\DateTime::__set_state(array(
'date' => '2014-12-23 23:41:10.000000',
'timezone_type' => 3,
'timezone' => 'Africa/Johannesburg',
)),
'impliedVolatility' => 12.201029873923984,
'inTheMoney' => true,
)),
7 =>
\EquityGPT\Results\OptionContract::__set_state(array(
'contractSymbol' => 'TFG.JOC0000000101',
'strike' => 8145.997827744658,
'currency' => 'Zac',
'lastPrice' => 8145.997827744658,
'change' => 0.0,
'percentChange' => 0.0,
'volume' => 38137251,
'openInterest' => 4,
'bid' => 0.0,
'ask' => 0.0,
'contractSize' => 'REGULAR',
'expiration' =>
\DateTime::__set_state(array(
'date' => '2015-12-23 23:41:10.000000',
'timezone_type' => 3,
'timezone' => 'Africa/Johannesburg',
)),
'lastTradeDate' =>
\DateTime::__set_state(array(
'date' => '2014-12-23 23:41:10.000000',
'timezone_type' => 3,
'timezone' => 'Africa/Johannesburg',
)),
'impliedVolatility' => 8.936861951476972,
'inTheMoney' => false,
)),
8 =>
\EquityGPT\Results\OptionContract::__set_state(array(
'contractSymbol' => 'TFG.JOC0000000102',
'strike' => 18378.705231944365,
'currency' => 'Zac',
'lastPrice' => 6126.235077314788,

'change' => 0.0,
'percentChange' => 0.0,
'volume' => 33914569,
'openInterest' => 9,
'bid' => 0.0,
'ask' => 0.0,
'contractSize' => 'REGULAR',
'expiration' =>
\DateTime::__set_state(array(
'date' => '2015-12-23 23:41:10.000000',
'timezone_type' => 3,
'timezone' => 'Africa/Johannesburg',
)),
'lastTradeDate' =>
\DateTime::__set_state(array(
'date' => '2014-12-23 23:41:10.000000',
'timezone_type' => 3,
'timezone' => 'Africa/Johannesburg',
)),
'impliedVolatility' => 14.851528366653218,
'inTheMoney' => false,
)),
9 =>
\EquityGPT\Results\OptionContract::__set_state(array(
'contractSymbol' => 'TFG.JOC0000000103',
'strike' => 8291.247704193278,
'currency' => 'Zac',
'lastPrice' => 8291.247704193278,
'change' => 0.0,
'percentChange' => 0.0,
'volume' => 46912950,
'openInterest' => 7,
'bid' => 0.0,
'ask' => 0.0,
'contractSize' => 'REGULAR',
'expiration' =>
\DateTime::__set_state(array(
'date' => '2015-12-23 23:41:10.000000',
'timezone_type' => 3,
'timezone' => 'Africa/Johannesburg',
)),
'lastTradeDate' =>
\DateTime::__set_state(array(
'date' => '2014-12-23 23:41:10.000000',
'timezone_type' => 3,
'timezone' => 'Africa/Johannesburg',
)),
'impliedVolatility' => 16.715246088586884,

'inTheMoney' => false,
)),
10 =>
\EquityGPT\Results\OptionContract::__set_state(array(
'contractSymbol' => 'TFG.JOC0000000104',
'strike' => 19495.185890502784,
'currency' => 'Zac',
'lastPrice' => 6498.395296834261,
'change' => 0.0,
'percentChange' => 0.0,
'volume' => 50696029,
'openInterest' => 6,
'bid' => 0.0,
'ask' => 0.0,
'contractSize' => 'REGULAR',
'expiration' =>
\DateTime::__set_state(array(
'date' => '2015-12-23 23:41:10.000000',
'timezone_type' => 3,
'timezone' => 'Africa/Johannesburg',
)),
'lastTradeDate' =>
\DateTime::__set_state(array(
'date' => '2014-12-23 23:41:10.000000',
'timezone_type' => 3,
'timezone' => 'Africa/Johannesburg',
)),
'impliedVolatility' => 15.719379926032762,
'inTheMoney' => false,
)),
11 =>
\EquityGPT\Results\OptionContract::__set_state(array(
'contractSymbol' => 'TFG.JOC0000000105',
'strike' => 12468.19708335196,
'currency' => 'Zac',
'lastPrice' => 6234.09854167598,
'change' => 0.0,
'percentChange' => 0.0,
'volume' => 47321810,
'openInterest' => 9,
'bid' => 0.0,
'ask' => 0.0,
'contractSize' => 'REGULAR',
'expiration' =>
\DateTime::__set_state(array(
'date' => '2015-12-23 23:41:10.000000',
'timezone_type' => 3,
'timezone' => 'Africa/Johannesburg',

```
  )),
   'lastTradeDate' =>
  \DateTime::__set_state(array(
     'date' => '2014-12-23 23:41:10.000000',
     'timezone_type' => 3,
     'timezone' => 'Africa/Johannesburg',
  )),
   'impliedVolatility' => 15.72043131799658,
   'inTheMoney' => false,
)),
12 =>
\EquityGPT\Results\OptionContract::__set_state(array(
   'contractSymbol' => 'TFG.JOC0000000200',
   'strike' => 6401.474050957961,
   'currency' => 'Zac',
   'lastPrice' => 6401.474050957961,
   'change' => 0.0,
   'percentChange' => 0.0,
   'volume' => 26156040,
   'openInterest' => 5,
   'bid' => 0.0,
   'ask' => 0.0,
   'contractSize' => 'REGULAR',
   'expiration' =>
  \DateTime::__set_state(array(
     'date' => '2015-12-23 23:41:10.000000',
     'timezone_type' => 3,
     'timezone' => 'Africa/Johannesburg',
  )),
   'lastTradeDate' =>
  \DateTime::__set_state(array(
     'date' => '2014-12-23 23:41:10.000000',
     'timezone_type' => 3,
     'timezone' => 'Africa/Johannesburg',
  )),
   'impliedVolatility' => 9.209502638703393,
   'inTheMoney' => false,
))
```