MSc in Finance & Banking (Part-Time)
Big Data & Machine Learning
FinTech Project

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

Team C

Week 1 – Portfolio Construction

Portfolio Structure

Defence Sector

The defence sector was selected as a key component of the portfolio, driven by geopolitical instability and the continuous rise in military expenditures. Companies in this sector provide stability through long-term government contracts and investments in advanced technologies such as artificial intelligence and cyberdefence.

Selected Companies

1. **Lockheed Martin (LMT)**: A global leader in aerospace and defence programs with a strong contract base, enhancing predictability and resilience.
2. **RTX Corporation (Raytheon)**: Offers a broad defence and aerospace portfolio, ensuring diversification and stability. Innovations in engines and missile systems further strengthen its competitiveness.
3. **Palantir Technologies (PLTR)**: Specializes in big data and AI solutions for military and security services. Its close cooperation with government agencies supports credibility and long-term growth.

Infrastructure Sector

The infrastructure sector represents a fundamental choice, as it is linked to public spending, the energy transition, and the ongoing need for network and construction upgrades. Companies in this sector are characterized by stability, diversification, and resilience to economic fluctuations.

Selected Companies

1. **Quanta Services (PWR)**: Focused on electrical networks and renewable energy projects, with a strategic role in the energy transition and major infrastructure works.
2. **Fluor Corporation (FLR)**: Operates internationally in large-scale engineering and construction projects. Its expertise in complex projects makes it a reliable partner for both public and private sectors.
3. **AECOM (ACM)**: Provides engineering, project management, and environmental solutions. Its broad project base supports diversification and long-term growth.

Technology Sector

The technology sector is the primary driver of innovation and growth, leading digital transformation, artificial intelligence, and data centers. Top companies combine advanced expertise, global reach, and diversified applications, strengthening portfolio dynamics.

Selected Companies

1. **NVIDIA Corporation (NVDA)**: A market leader in GPUs and computing power solutions, playing a pivotal role in AI development. Continuous innovation underpins its dominance in the data center market.
2. **Advanced Micro Devices (AMD)**: Active in CPUs, GPUs, and cloud/AI infrastructure solutions, steadily expanding its portfolio and reinforcing its competitive position.
3. **QUALCOMM Incorporated (QCOM)**: Specializes in wireless communication technologies and 5G chipsets. Its strategic position in telecommunications ensures long-term prospects.

Metals Sector

The metals sector offers strategic diversification, as it is influenced more by commodity prices than the economic cycle. Metals are also essential for the energy transition and infrastructure, enhancing the sector's significance.

Selected Companies

1. **Alcoa Corporation (AA)**: One of the largest aluminum and bauxite producers globally. Its strong raw material position supports long-term demand.
2. **Reliance Steel & Aluminum Co. (RS)**: A leading distributor and processor of metals with a diversified product portfolio. Its broad client base enhances stability and resilience.
3. **MDU Resources Group, Inc. (MDU)**: Combines energy, construction, and utility services in a diversified model. Its regulated activities provide predictability and steady cash flows.

Company	Sector	P/E	FCF/Equity	EV/EBITDA	D/E	Net Profit Margin
Lockheed Martin (LMT)	Defence	26,6	83,5%	17,3	4	7,0%
RTX Corporation (Raytheon)	Defence	34,8	6,5%	17,5	0,7	7,4%
Palantir Technologies (PLTR)	Defence	563,2	21,4%	663,9	0,04	16,3%
Quanta Services (PWR)	Infrastructure	58,5	20,2%	26,4	0,65	4,0%
Fluor Corporation (FLR)	Infrastructure	1,7	16,8%	13,3	0,18	25,0%
AECOM (ACM)	Infrastructure	25,35	32,4%	14,5	1,2	3,5%
NVIDIA Corporation (NVDA)	Technology	50,66	90,7%	41,53	0,11	52,4%
Advanced Micro Devices (AMD)	Technology	27,03	7%	46,15	0,065	9,6%
QUALCOMM Incorporated (QCOM)	Technology	15,36	44,2%	12,33	0,54	26,8%
Alcoa Corporation (AA)	Metals	8,01	0,8%	4,36	0,41	7,9%
Reliance Steel & Aluminum Co. (RS)	Metals	21	13,8%	12,6	0,24	5,4%
MDU Resources Group, Inc. (MDU)	Metals	18	-0,8%	5,7	0,8	11,6%

Crypto Sector

Three leading pillars of the crypto market were chosen, each covering distinct investment and technological profiles: Bitcoin (Store of Value), Ethereum (Web3/DeFi infrastructure), and Solana (high-performance Layer-1 for large-scale applications).

Selected Cryptos

1. **Bitcoin (BTC)**: “Digital gold” with scarcity (21M supply), high liquidity, institutional adoption, and strongest security (PoW). Acts as a long-term store of value and potential hedge.
2. **Ethereum (ETH)**: The leading smart contract platform with >60% share of DeFi TVL. Transition to Proof-of-Stake reduces energy use >99%, while Layer-2 scaling supports adoption. A mature Web3/DeFi infrastructure.
3. **Solana (SOL)**: A high-performance blockchain with very low transaction fees, innovative Proof-of-History mechanism, and strong ecosystem growth (DeFi, NFTs, Web3 gaming)

Conclusion

BTC serves as a foundational store of value, ETH as infrastructure for Web3/DeFi, and SOL as an innovative high-throughput chain. Together, they enhance portfolio diversification across value storage, infrastructure, and innovation.

Investor Persona

Our portfolio aligns with an Aggressive Investor Persona, designed to capture high-growth opportunities and benefit from global economic cycles. The allocation combines innovation leaders in technology (NVDA, AMD, QCOM, PLTR) with industrial and infrastructure firms (ACM, FLR, PWR, MDU) strategically positioned to benefit from long-term investment in energy and construction. Exposure to basic materials (AA, RS) adds leverage to commodity upswings, while defence holdings (RTX, LMT) provide resilience. Cryptocurrencies further diversify the portfolio and offer asymmetric return potential. The mix is deliberately growth-oriented, aimed at investors with a long horizon and high risk tolerance, seeking superior returns through a diversified blend of cyclical, innovative, and resilient sectors.

Macroeconomic Indicators

1. **GDP Growth**:
    - Strong GDP growth signals expanding corporate earnings and higher demand, boosting risk assets; weak or negative growth raises recession fears and pressures sentiment.
2. **Inflation Rate**:
    - Rising inflation reduces purchasing power and often triggers tighter monetary policy, weighing on valuations; stable or falling inflation supports markets.
3. **Unemployment Rate**:
    - Low unemployment reflects economic strength and higher consumption but may add wage inflation; high unemployment signals weakness and hurts markets.
4. **Interest Rate (Monetary Policy)**:
    - Higher rates increase borrowing costs and restrict liquidity, slowing growth and risk assets; lower rates support growth and markets.
5. **Consumer Confidence**:
    - Rising confidence supports consumption expectations and market sentiment; falling confidence signals caution and potential demand weakness.
6. **Retail Sales**:
    - Strong sales indicate healthy demand and economic momentum; weak sales raise concerns about slowdown.
7. **Industrial Production**:
    - Growth in production reflects strong manufacturing and economic expansion; contraction signals weakening demand.
8. **Housing Starts**:
    - Rising housing starts show real estate strength and future economic activity; declines indicate slowdown and potential weakness.

Week 2 – Data Processing & Imputation

Introduction

We integrated equity and cryptocurrency data to explore imputation techniques, evaluate utility functions, and analyze portfolio risk–return characteristics. By combining twelve U.S. equities (defense, infrastructure, technology, metals) with cryptocurrencies (BTC, ETH, SOL), the dataset was aligned on a seven-day calendar. Each method was assessed in the context of weekend stock gaps, ensuring consistency between equities and continuously trading cryptos. The analysis connected technical methods with investor philosophy, reflecting an Aggressive Investor Persona willing to accept volatility for asymmetric growth potential.

Weekend Simple Imputation

The first step was to handle weekend gaps in the stock data to match the seven-day cycle of cryptocurrencies. Simple mean and median imputations were applied to fill missing Saturday and Sunday closes. Mean imputation preserved the overall average (≈140.69) and slightly reduced variance, while median imputation increased the mean marginally (≈141.11) with similar variance.

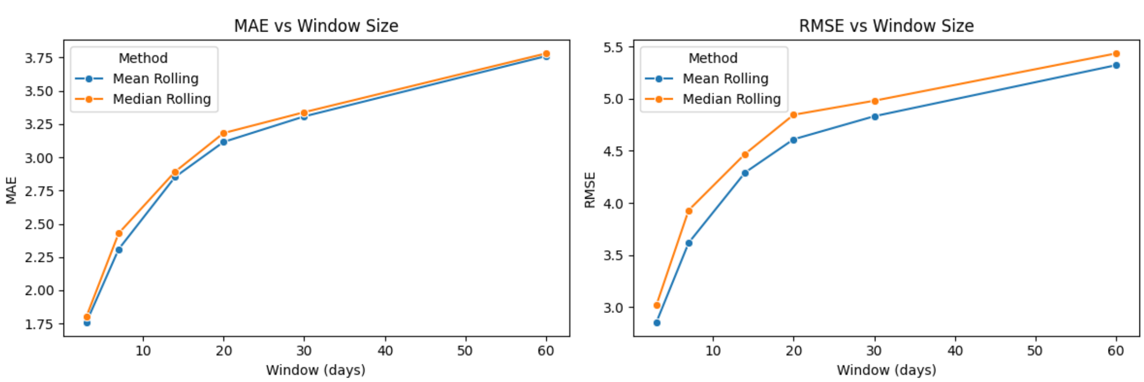
Although both methods provided a complete dataset with no weekend NaNs, they differ in robustness. Mean imputation is sensitive to outliers, which makes it more suitable for aggressive investors seeking to capture volatility signals. Median imputation is more resilient to extreme values, a safer choice for risk-averse investors as it dampens sharp spikes. For our aggressive portfolio, mean imputation was preferred since it preserves large swings critical for analyzing technology and crypto assets, while median remains a useful fallback for more conservative segments such as defense stocks.

Rolling Window Imputation

Rolling averages with window sizes of 3, 7, 14, 20, 30, and 60 days were tested against interpolated weekend reference values. This allowed us to measure how well each method captured short-term dynamics while filling the gaps.

The results confirmed that short windows are most effective. The 3-day rolling mean produced the lowest errors (MAE ≈ 1.76, RMSE ≈ 2.86), closely followed by the 3-day rolling median (MAE ≈ 1.80, RMSE ≈ 3.02). Expanding the window diluted responsiveness—already at seven days, errors exceeded 2.3, and windows longer than two weeks pushed RMSE above 4.6. This shows that longer windows smooth fluctuations but at the cost of missing important short-term movements.

This trade-off reflects different investor philosophies. Aggressive investors benefit from shorter windows that capture recent volatility, while risk-averse investors might prefer the stability of longer windows, even if accuracy suffers.



Evaluation of Imputation Metrics

To better understand the accuracy of different strategies, we assessed weekend fills with MAE, RMSE, MAPE, and sMAPE. Using the interpolated weekend values as a benchmark made the contrast between methods clear.

Mean and median replacements produced high errors (MAE > 12, RMSE ≈ 18), confirming their inability to preserve local price dynamics. By contrast, the hybrid forward/backward fill approach delivered exceptional accuracy, with MAE ≈ 0.40 and RMSE ≈ 0.66. Both MAPE and sMAPE dropped below 0.3%, showing that the hybrid method effectively reconstructed weekend behavior with minimal distortion.

The advantage of hybrid fill lies in structure: by combining forward-fill, backward-fill, and column mean fallback, it preserves short-term continuity and avoids the flat-lining effect of simple statistics. RMSE highlighted tail risks, MAE measured average deviations in price levels, and percentage-based metrics offered scale-free interpretability. Together, these results confirmed hybrid fill as a strong baseline for weekend imputation.

Method	**MAE**	**RMSE**	**MAPE (%)**	**sMAPE (%)**
Hybrid Forward/Backward Fill	0.3972	0.6618	0.2694	0.2694
Simple Median	12.6423	18.5098	11.8923	10.8596
Simple Mean	12.9823	17.7981	11.8954	11.1918

Interpolation Approaches

We also implemented linear, quadratic, cubic, and simple exponential smoothing to test their performance on weekend gaps. Among these, linear interpolation emerged as the most accurate.

When evaluated against the interpolated weekend reference, linear interpolation essentially matched the true curve, with errors approaching zero (MAE ≈ 0, RMSE ≈ 0). Quadratic and cubic fits performed reasonably well but introduced slight oscillations, leading to MAE values between 1.2–1.3 and RMSE around 2.1–2.3. Simple exponential smoothing with $\alpha = 0.3$ lagged during sharp Friday-to-Monday shifts, producing RMSE above 3.

The mechanics explain these differences. Linear interpolation connects Friday and Monday prices with a straight line, making it ideal for the two-day weekend gap. Polynomial fits can capture multi-day drifts but risk instability when gaps are short. Exponential smoothing stabilizes values but cannot anticipate sudden changes after a non-trading period.

Given these results, linear interpolation stands out as the preferred default for weekend alignment, combining accuracy, efficiency, and robustness.

Handling Numerical Pathologies

To ensure the reliability of our results, we implemented safeguards against numerical issues. A denominator floor combined with sMAPE eliminated infinite or undefined values in percentage-based metrics. This prevented distortions from near-zero denominators and ensured all reported metrics were finite and consistent.

Additionally, correlations recomputed after linear interpolation remained within the valid range of [-1, 1]. A clipping safeguard was introduced to protect against distortions if a near-constant series appeared in future runs. Together, these guardrails ensured the stability and reproducibility of our analysis.

Advanced Imputation Techniques

We further explored advanced methods such as KNN and MICE to test whether model-based approaches could outperform simpler routines.

KNN with three neighbors performed strongly, reducing MAE to ≈2.88 and RMSE to ≈4.48—substantially better than mean/median fills and close to the handcrafted hybrid. Increasing k to 5 or 10 smoothed reconstructions but added bias, raising MAE into the 2.9–3.1 range.

MICE with a RandomForest regressor achieved similar results (MAE ≈ 3.01, RMSE ≈ 4.67), while the BayesianRidge variant struggled with sharp weekend movements, producing higher errors (MAE ≈ 5.28). Overall, KNN with a small k was the most practical upgrade over simple fills, while RandomForest MICE remains a valuable fallback when more complex, model-driven imputations are desired.

Utility Functions and Investor Behavior

Utility functions capture how investors perceive risk and reward. Three profiles were considered: risk-averse U(W)=sqrt{W}, risk-neutral U(W)=W, and risk-seeking U(W)=W^2. Risk-averse investors prioritize capital preservation, gaining less satisfaction from each additional unit of wealth while being highly sensitive to losses. They are drawn to stable equities with predictable revenues, such as LMT and RTX. Risk-neutral investors view outcomes linearly, focusing on expected returns regardless of variance. This leads to balanced portfolios that include infrastructure firms like FLR, ACM, and PWR. Risk-seeking investors place greater weight on upside potential, favoring high-growth assets such as NVDA, PLTR, and cryptocurrencies.

Our portfolio embodies a risk-seeking approach, combining stable anchors in defense with high-volatility technology and crypto assets to maximize long-term upside. This reflects the Aggressive Investor Persona. Importantly, the choice of utility function also influences data handling. Risk-averse investors would apply conservative imputations, such as median fills, to avoid overstating returns. Risk-neutral investors might favor mean or regression-based methods, while risk-seeking investors may extrapolate missing values across correlated assets, such as projecting Ethereum returns from Bitcoin or Solana trends.

For example, if ETH data were missing for a week, a risk-averse investor would exclude it, a neutral investor would average it, and a risk-seeker would project based on related crypto movements. This framework shows that utility functions guide both portfolio construction and technical decisions in data management. By adopting $U(W)=W^2$, our strategy captures why aggressive investors embrace volatility, tolerate noise in imputations, and pursue asymmetric returns. Defense assets provide a stabilizing base, but the portfolio's core rests on risk-seeking behavior in technology and crypto.

Final Dataset and Portfolio Returns

To construct the final dataset, we filled weekend stock gaps using linear interpolation. This method was chosen after systematic benchmarking and Monte Carlo–style evaluation, where thousands of weekend-like intervals were masked and re-imputed across multiple runs. The repeated tests confirmed that linear interpolation consistently delivered the lowest errors (MAE ≈ 0, RMSE ≈ 0), outperforming alternatives such as hybrid forward/backward fill, quadratic and cubic fits, and exponential smoothing.

Linear interpolation was particularly well-suited to the weekend alignment problem because weekend gaps are always exactly two consecutive days, bounded by reliable Friday and Monday prices. This makes a straight-line connection the most accurate and efficient solution. The Monte Carlo evaluation reinforced the robustness of this choice, demonstrating that the method not only performed best on average but also maintained stability across resampled trials.

With the gaps filled, portfolio returns were computed in both simple and log forms, with log returns offering more stable and consistent distributions for risk analysis. Portfolio evaluation showed a positive Sharpe ratio despite high variance, underscoring the high-risk, high-return nature of the Aggressive Investor Persona. Volatility was driven primarily by crypto and technology exposures, while defense and infrastructure holdings provided diversification that mitigated some of the risk.

This outcome illustrates the fundamental risk–return trade-off: cryptocurrencies introduce volatility and speculative upside, while equities in defense, infrastructure, and industrials provide stability. By explicitly selecting linear interpolation, validated through Monte Carlo benchmarking, the dataset was made robust and reliable for downstream analysis, ensuring that both the technical foundation and the investor philosophy were aligned.

Conclusion

This stage of the project demonstrated the close link between technical imputation choices and investor philosophy. Weekend-specific imputations confirmed that simple statistical fills were inadequate, hybrid forward/backward provided a robust baseline, and linear interpolation emerged as the most accurate and efficient solution for aligning stock and crypto calendars.

The decision to adopt linear interpolation was not based on a single test but on systematic Monte Carlo evaluation. By repeatedly masking weekend-like intervals and re-imputing them, we validated that linear interpolation consistently delivered near-zero error and maintained stability across resampled trials. This evidence-based approach ensured that our chosen method was not only theoretically suitable for the two-day weekend gap but also empirically the most reliable.

Advanced model-driven techniques such as KNN and MICE provided incremental gains in some contexts, but with added complexity and without surpassing linear interpolation for the weekend problem.

Ultimately, these results highlight that imputation is not merely a technical detail but a reflection of the broader investment philosophy. By adopting linear interpolation, validated through Monte Carlo benchmarking, we reinforced the Aggressive Investor Persona: embracing volatility and precision where it matters most, while maintaining a robust foundation for portfolio construction and risk–return analysis.