Risk Analytics Pipeline – VaR Calculation & Reporting

Miloš Novaković

December 12, 2025

Abstract

This project develops a modular, end-to-end Risk Analytics Pipeline designed to showcase data engineering, data science, and financial risk-management skills through the calculation of historical Value-at-Risk (VaR). The pipeline ingests multi-source (different business lines) position data and historical market prices, stores them in a documented relational schema, and performs data cleaning using mainly python pandas and SQL-based transformations. A 1-day, 99% historical VaR with a one-year lookback window is computed at multiple aggregation levels, including individual business lines, metal categories, and the whole consolidated portfolio. The system emphasizes transparency through clear assumptions, reproducible methodology, and modular code design. Final outputs include a structured Excel report summarizing VaR results, along with a technical markdown documenting the approach, challenges, and recommendations.

1 Introduction - Data Ingestion

We began by ingesting two primary datasets: a position dataset and a historical price dataset, both containing exclusively futures contract information. The position data originated from three distinct business lines—copper, proprietary trading, and lead/zinc. Each business line may contain an arbitrary number of trading strategies, classified into one of three categories: directional (flat-price exposure, expressing a view on outright price movements) and calendar spread (same exchange, different maturities, capturing views on contango or backwardation), [1] and [3]), as well as arbitrage (same maturity, different exchanges) [2]. A DuckDB database was implemented to store and structure the ingested datasets, with the corresponding schema provided in the repository : git.

2 Data Processing and Cleaning

All available price series Copper LME [4], Copper CME [5], and Zinc LME [6] were plotted across their respective maturities to visually inspect potential outliers. Several apparent anomalies identified during the visual review were subsequently verified manually and confirmed to be accurate. Descriptive statistics were computed for each maturity (including count, minimum, maximum, mean, standard deviation, and 25/50/75 percentiles), and no issues were identified that would warrant concern. While the current analysis did not reveal problematic data points, there is significant opportunity to extend this step with more advanced outlier-detection techniques. Additionally, due to the contract structure of LME futures—expiring on the Monday preceding the third Wednesday—certain maturities naturally contain shorter time series, particularly for contracts that expired earlier within the fixed price-data window (1 October 2024 to 30 October 2025). This behavior is expected. In such cases, a zero-order polynomial interpolation (constant fill) was applied to align maturities for VaR calculations. Following examples of price time series.

3 Risk Calculation

A historical VaR calculation (1 day horizon, 99% confidence level, 1 year lookback) was implemented for each business line (Copper [7] and [8]; Prop [9] and [10]; Zinc and Lead [11] and [12]), each metal (Copper [13] and [14]; Zinc [15] and [16]), and for the consolidated portfolio (i.e., all three business lines combined [17] and [18]). The one-year lookback window was defined as 260 trading days, based on 52 trading weeks of five business days each, resulting in a price time series of length 260 at a daily frequency. As noted previously, certain contracts exhibit only partial data availability. For example, for the Copper COMEX December 2025 maturity, if VaR is calculated as of 1 September 2025, the most recent available price is from 1 October 2025. Consequently, only 226 price points are available between 1 September 2025 (most recent) and 1 October 2024 (earliest). To obtain the full 260-day history, the remaining 34 observations (260 less 226) were backfilled using a zero-order polynomial (constant) interpolation, carrying the earliest available price backward. That is why in our example we choose to calculate VaR as of 1st of September 2025. Although the appropriateness of this approach can be debated, it is reasonable in this context: futures prices, modeled as martingales, imply that the best unbiased estimate of future values is the current expectation, supporting the use of a constant backfill for missing historical observations. This approach was applied consistently across all business lines (Proprietary Trading, Copper, and Zinc), across metals, and at the aggregated portfolio level. After aligning price series, we converted prices into standardized units (USD/MT or USD/lb for CME contracts) and computed daily price changes, reflecting MTM settlement of futures. Multiplying these daily changes by the corresponding position sizes (in metric tons or pounds) produced a 259-observation time series of daily PnL. The historical 99% VaR was then calculated by taking the 1st percentile of this distribution. The resulting PnL distributions and the corresponding VaR thresholds (highlighted as red dashed lines) are shown in the charts that accompany this section.

4 Results

The Results are summed up in the Table [1].

The Proprietary Trading business line carries both directional exposure (e.g., long copper) and spread exposure (e.g., borrowing copper and being long backwardation). From a theoretical risk perspective, directional positions typically exhibit significantly higher volatility and therefore higher VaR than pure spread positions. Consistent with this, we expected the Proprietary Trading book to display greater risk than the Zinc/Lead business line, which holds exclusively spread positions. Additionally, the unit VaR of copper is structurally higher than that of zinc or lead, further reinforcing this expectation.

An unexpected result emerged, however: the Copper business line, which holds only a long arbitrage position (long LME, short CME, 6kt Dec-2025), displayed a materially higher VaR than the Proprietary Trading book with outright directional risk. This outcome is anomalous under normal market conditions, as LME and CME copper prices are typically strongly positively correlated. Consequently, an arbitrage position involving equal and opposite exposures on both exchanges usually carries far less risk than an equivalent outright directional position.

This anomaly is explained by a significant correlation breakdown that occurred over the past year. The most notable dislocation occurred on 31 July 2025, when U.S. President Donald J. Trump announced that there would be no tariff increase on domestic U.S. copper (i.e., CME deliverable copper). Prior to the announcement, markets had priced in an expected tariff of roughly +30%. Following the announcement, CME copper prices sold off sharply declining approximately USD 2,600 USD/MT day-on-day while LME copper remained relatively stable. As a result, the CME-minus-LME spread compressed dramatically, and the historically strong positive correlation between the two exchanges weakened substantially.

From a risk perspective, this meant that the short CME leg of the arbitrage position no longer provided the expected hedge against the long LME leg. Instead, it contributed additional risk, as reflected in the elevated historical VaR. The VaR results therefore accurately capture this market dislocation and the temporary breakdown in cross-exchange pricing relationships.

It is also important to highlight the effect of diversification. When calculating the aggregated Copper VaR, the result is approximately USD 2.76 million, which is lower than the sum of the individual Copper-related VaRs from the Proprietary Trading business line (USD 0.20 million) and the Physical

Copper business line (USD 2.96 million). This demonstrates a clear diversification benefit, driven by less-than-perfect correlation between the underlying exposures.

Portfolio grouping	Hist. VaR 99% (USD)
BL Copper	2'955'973
BL Prop	197'331
BL Zinc/Lead	51'000
Metal Copper	2'759'541
Metal Zinc	51'000
Total	2'769'598

Table 1: Calculated Historical VaR on requested groups.

5 Supporting Plots

The following are supporting plots and figures.

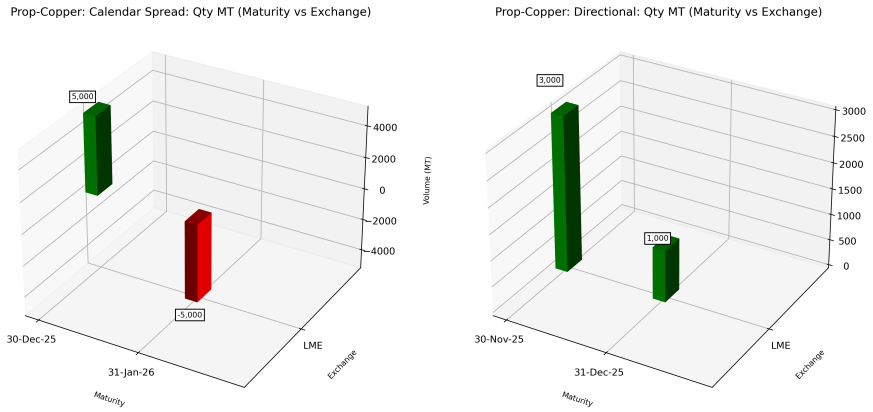

Figure 1: Business line Proprietery trading: Directional and Spread Copper strategy positioning

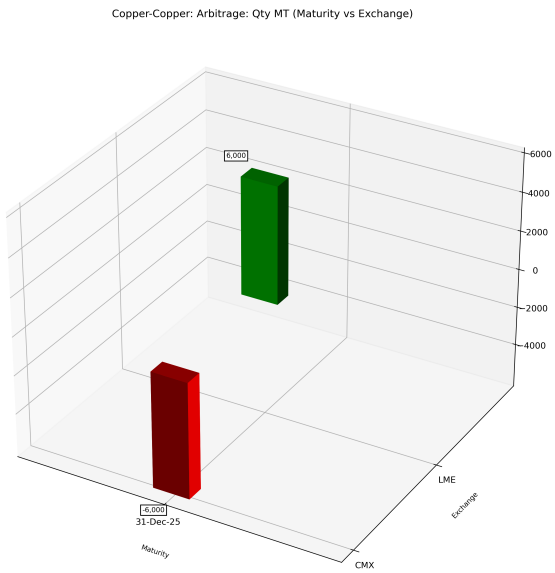

Figure 2: Business line Copper: LME/CME Copper Arbitrage strategy positioning

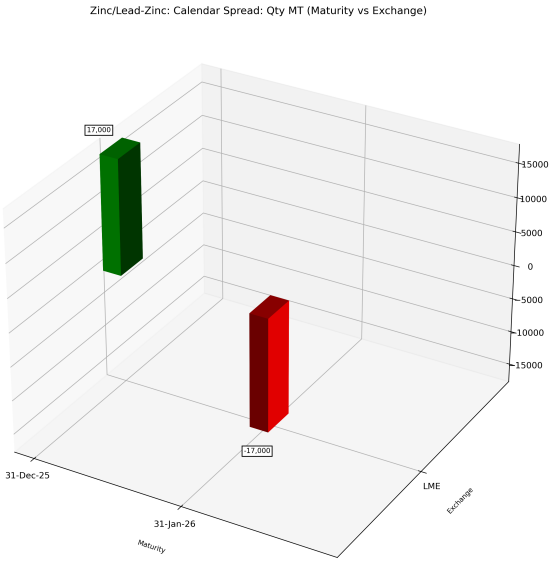

Figure 3: Business line Zinc/Lead: Calendar Spread Zinc strategy positioning

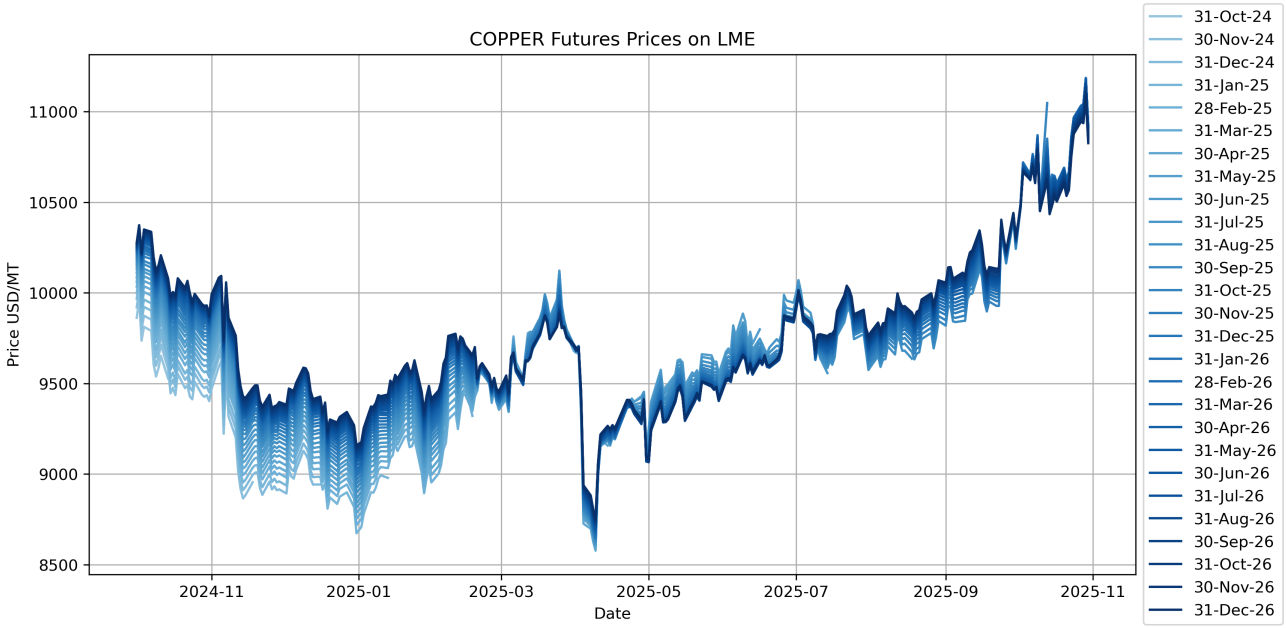

Figure 4: Copper LME historical prices across maturities

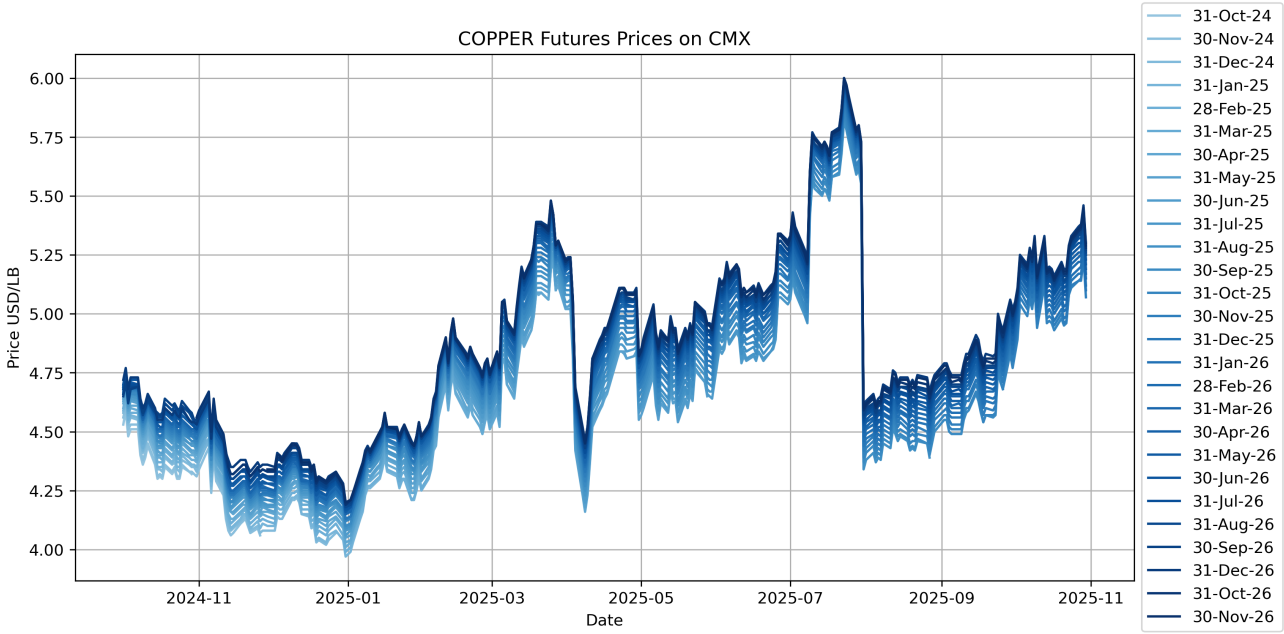

Figure 5: Copper CME historical prices across maturities

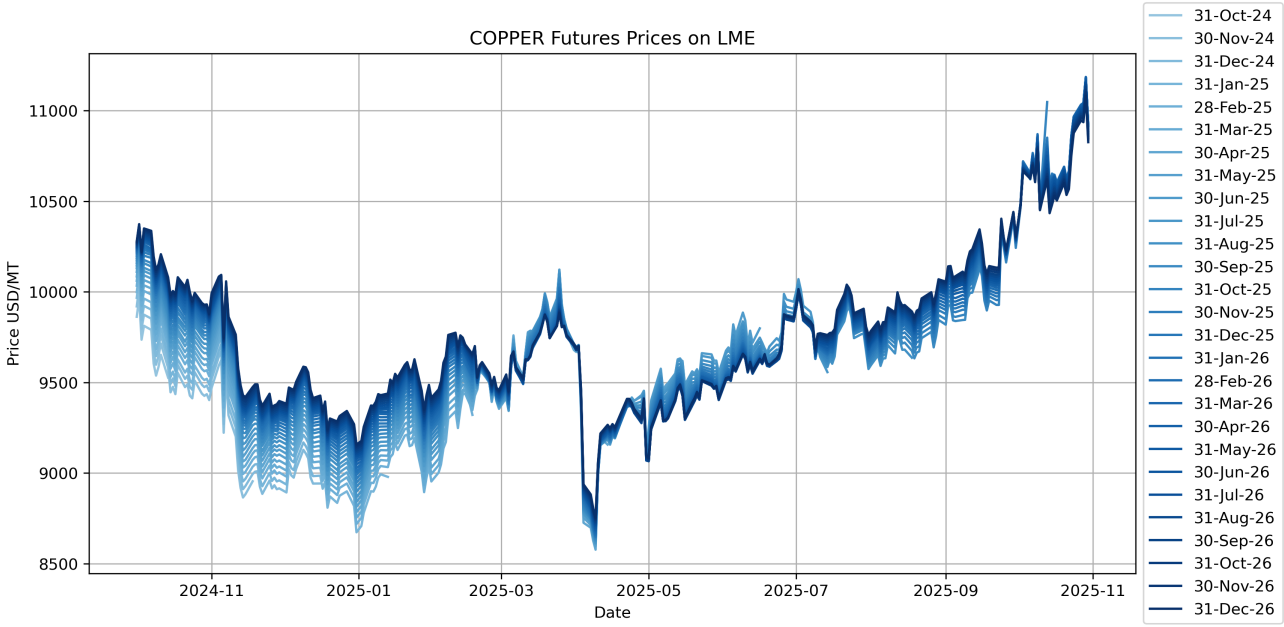

Figure 6: Zinc LME historical prices across maturities

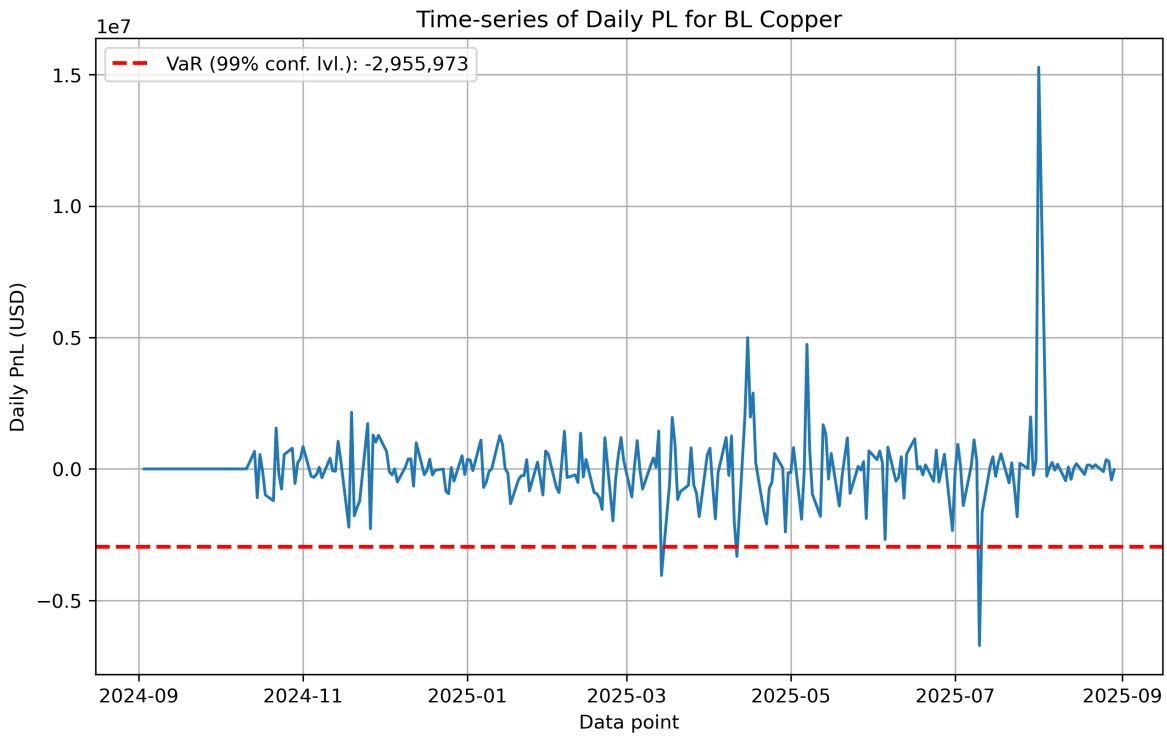

Figure 7: Business Line Copper PnL

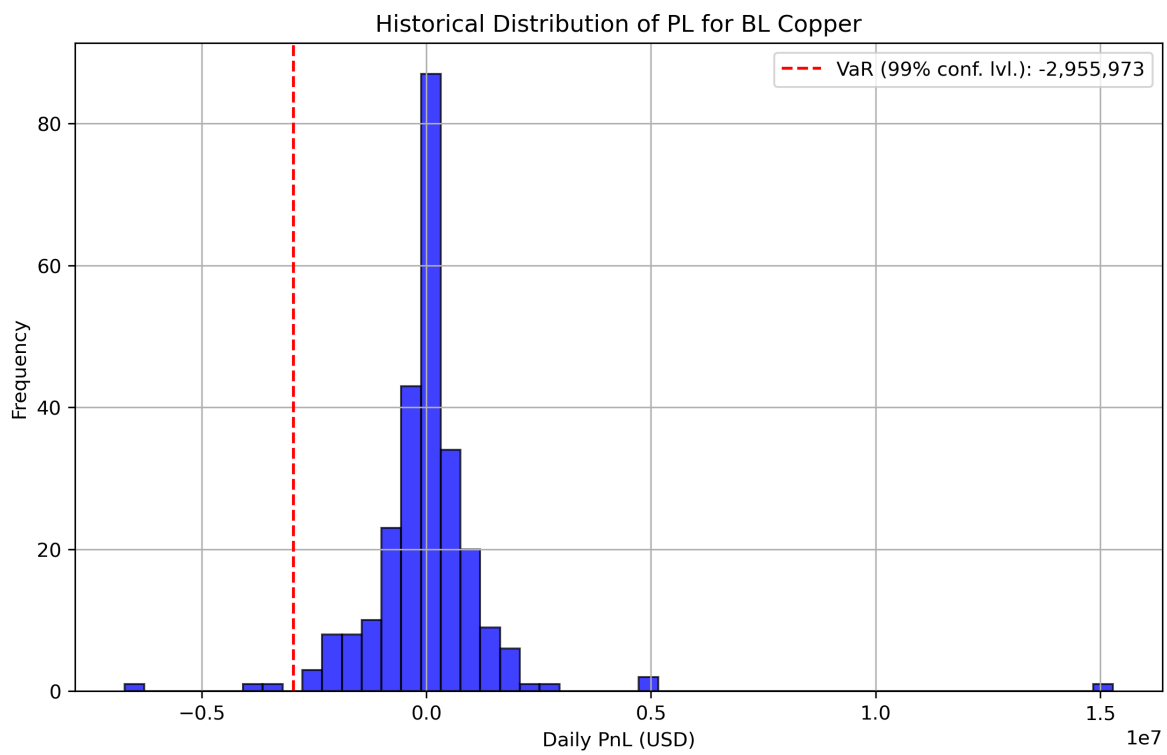

Figure 8: Business Line Copper PnL Distribution

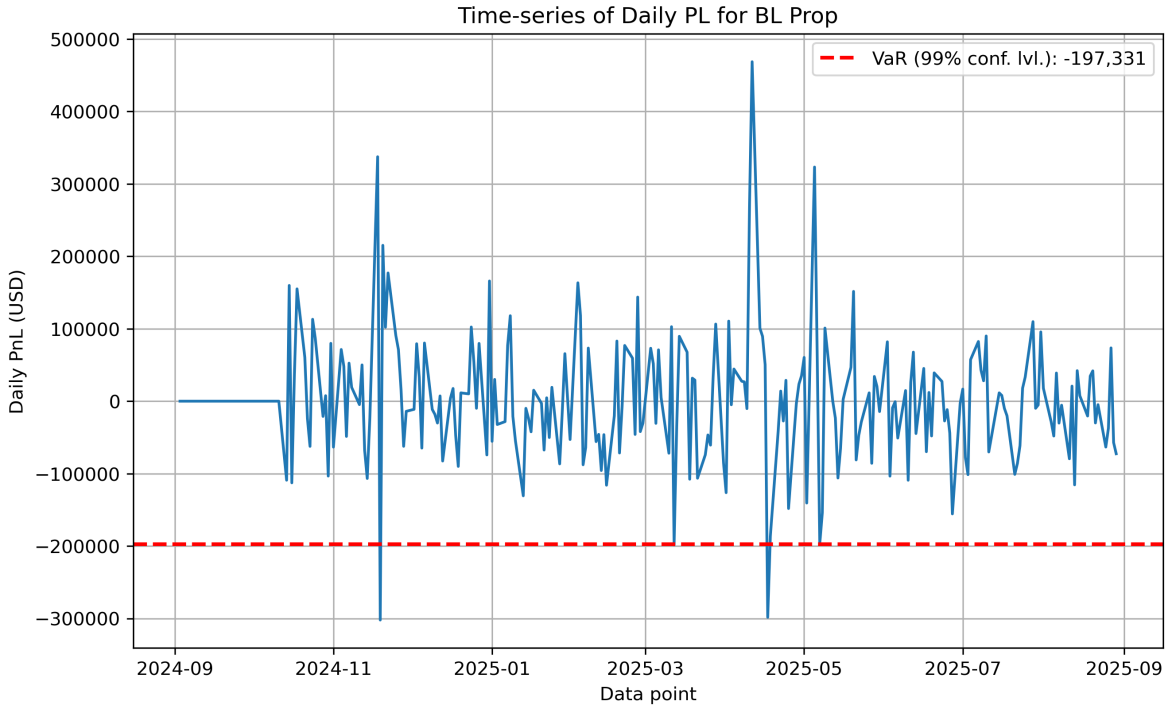

Figure 9: Business Line Prop PnL

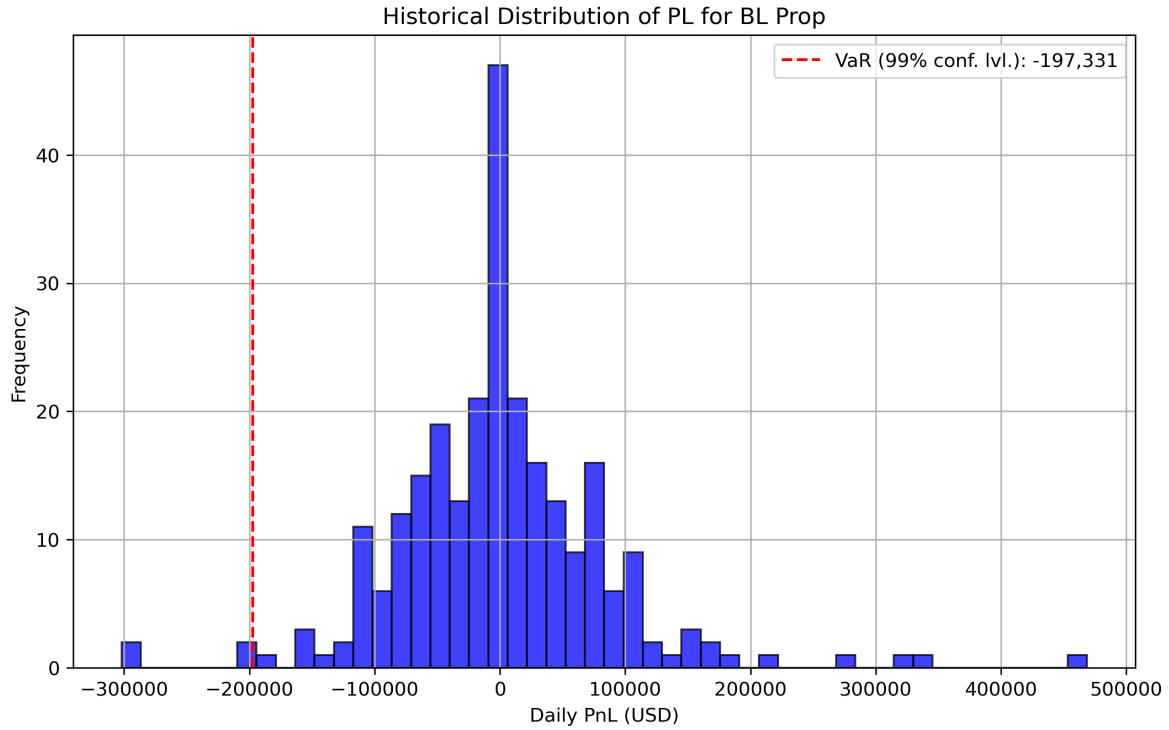

Figure 10: Business Line Prop PnL Distribution

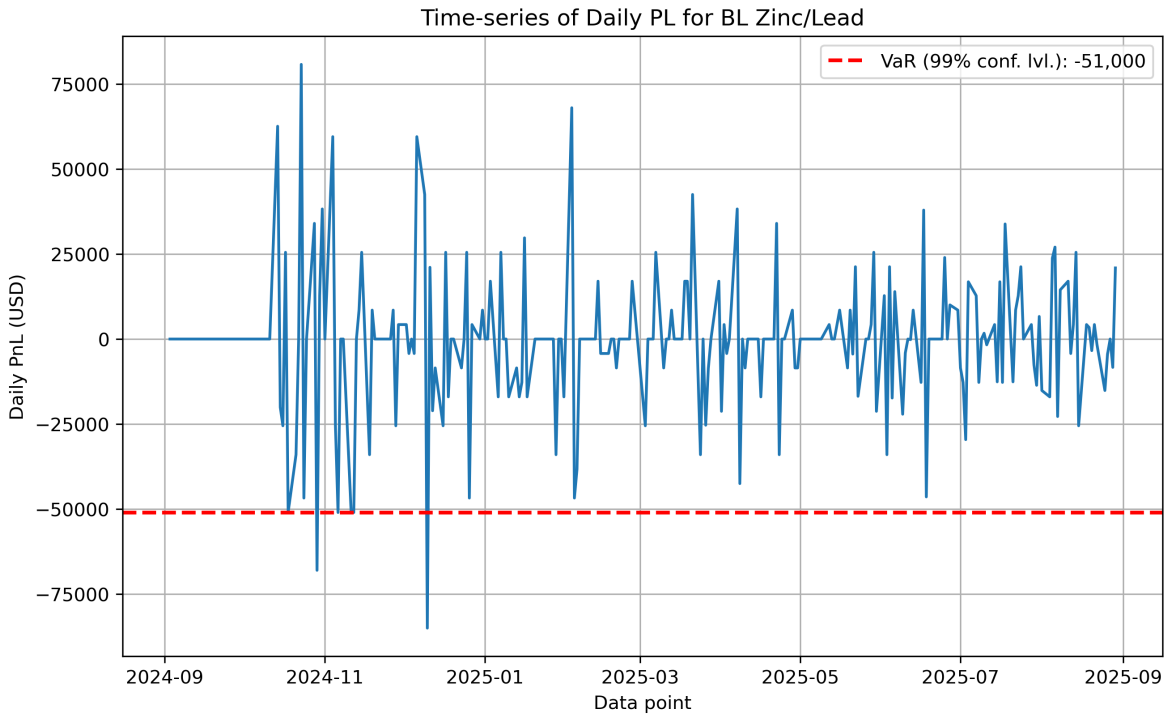

Figure 11: Business Line Zinc and Lead PnL

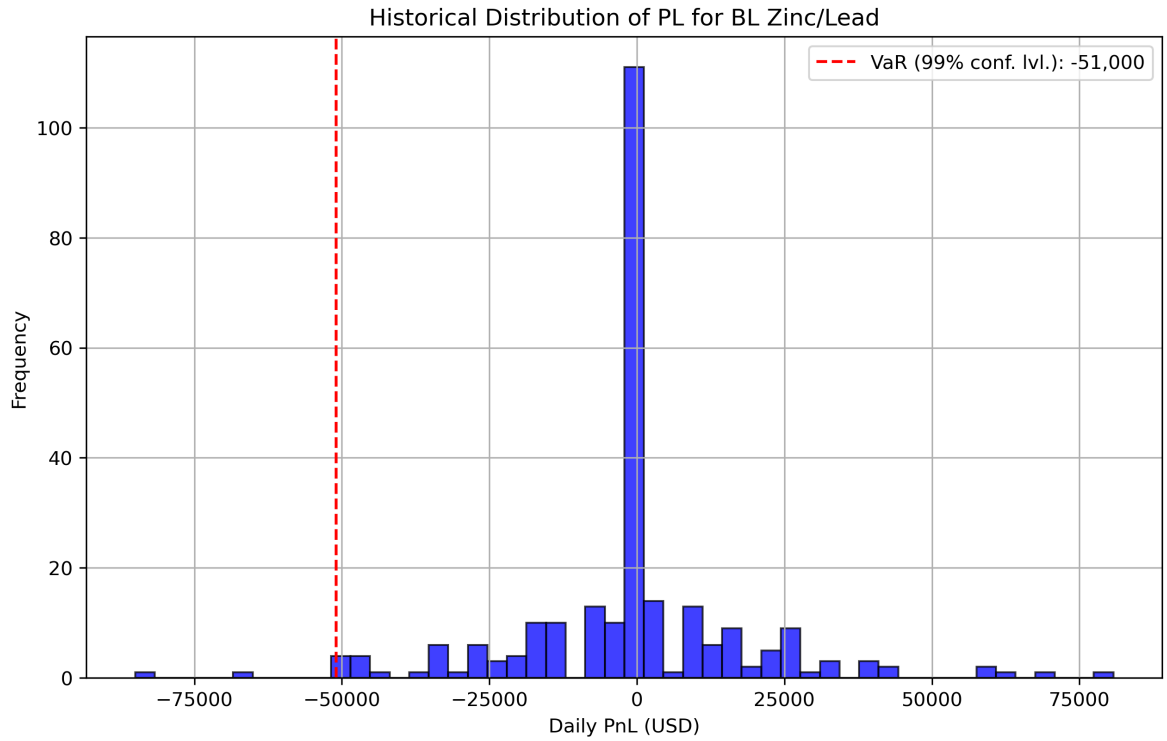

Figure 12: Business Line Zinc and Lead PnL Distribution

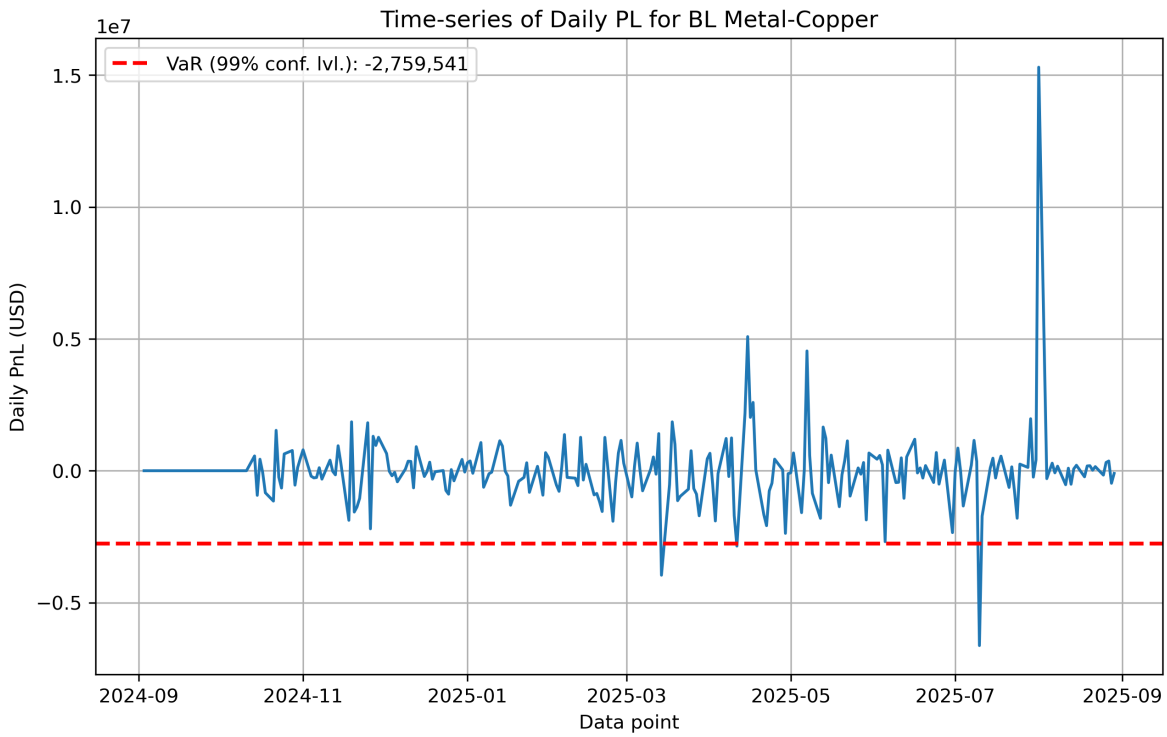

Figure 13: Across all Copper PnL

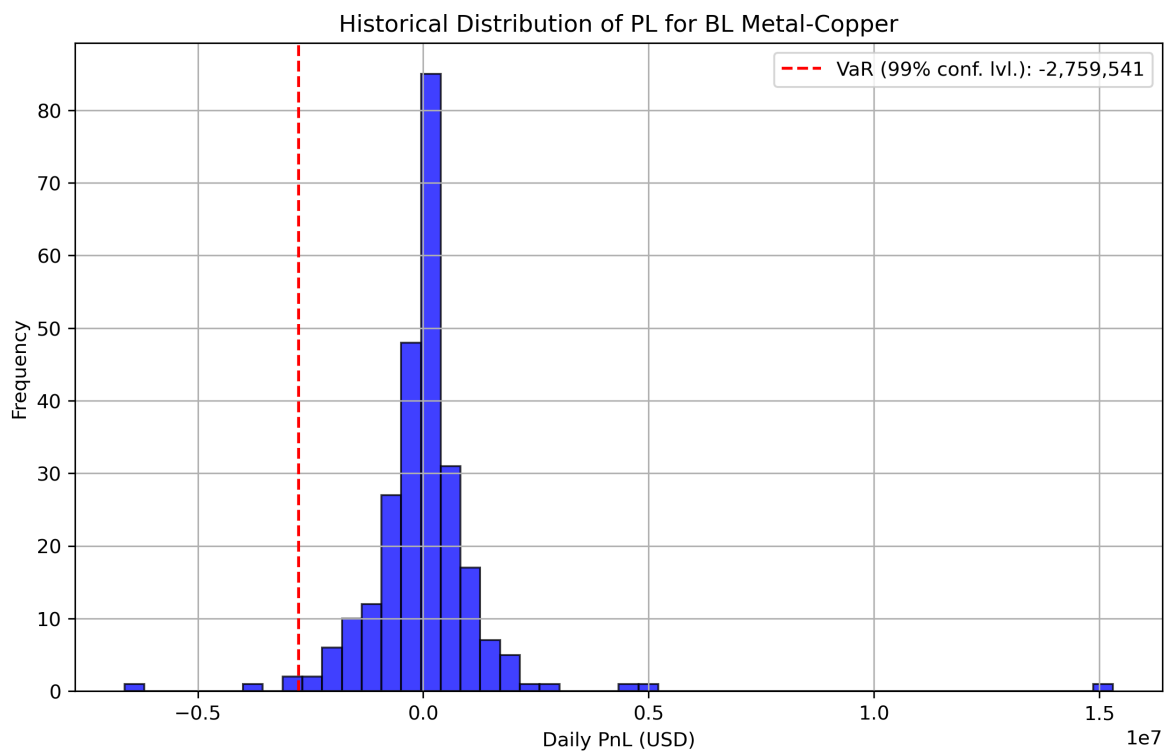

Figure 14: Across all Copper PnL Distribution

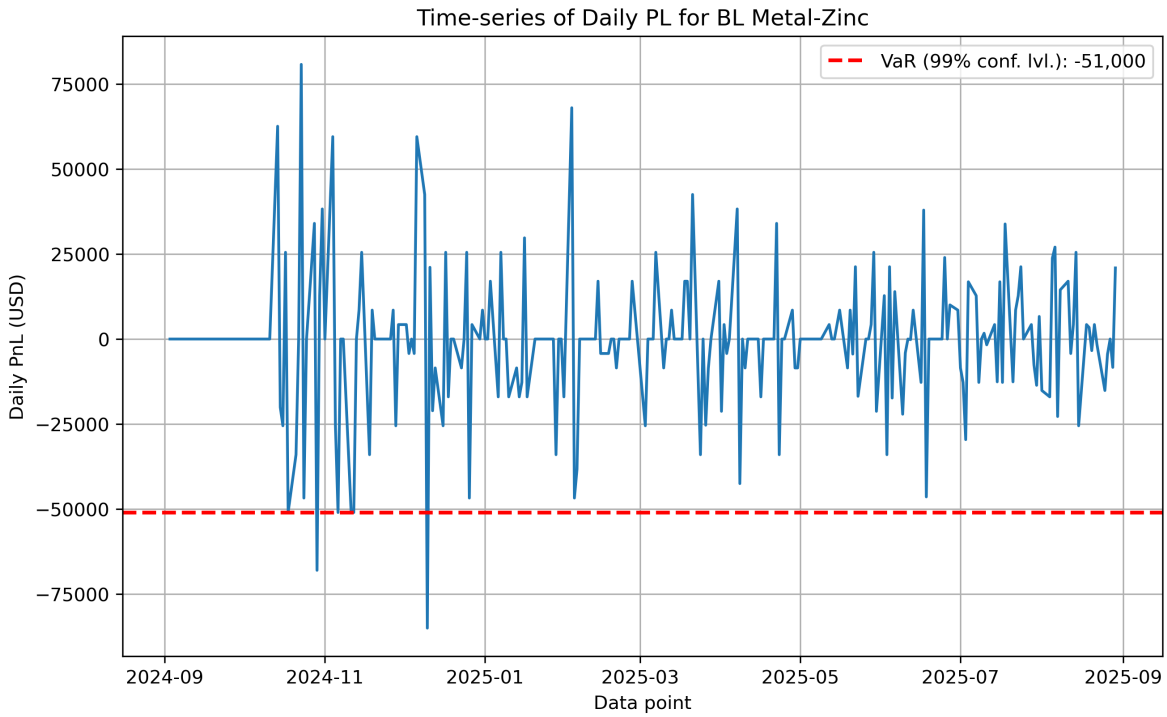

Figure 15: Across all Zinc PnL

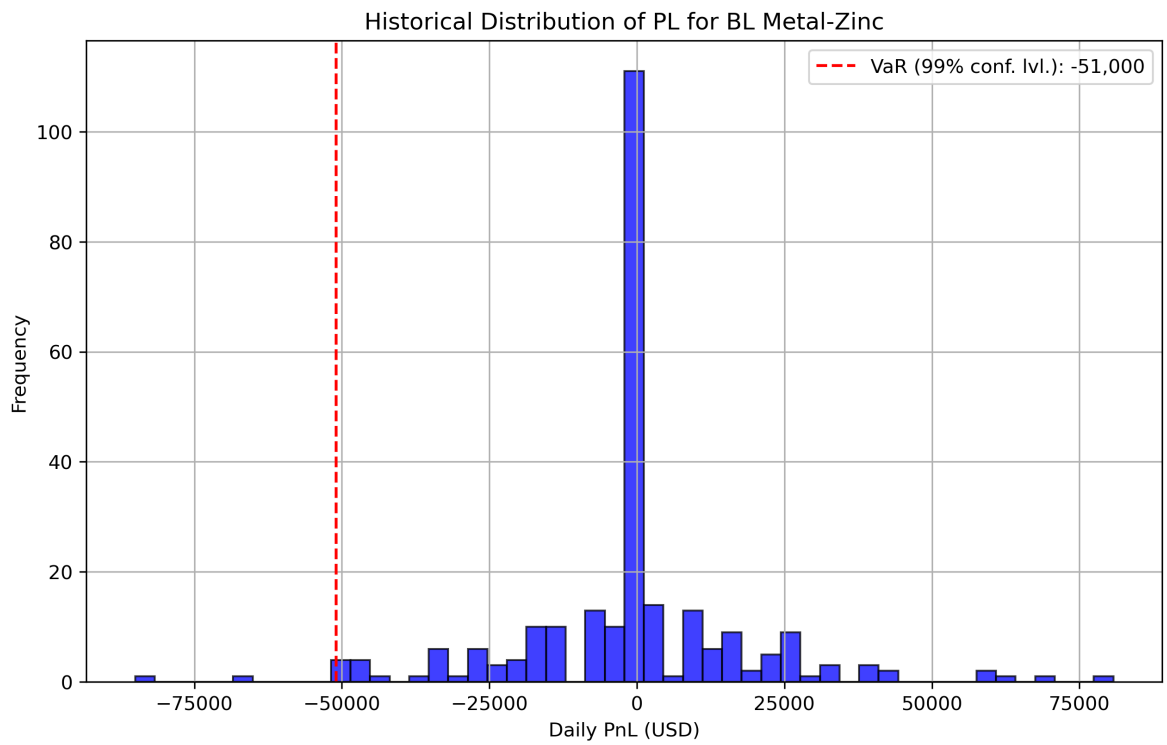

Figure 16: Across all Zinc PnL Distribution

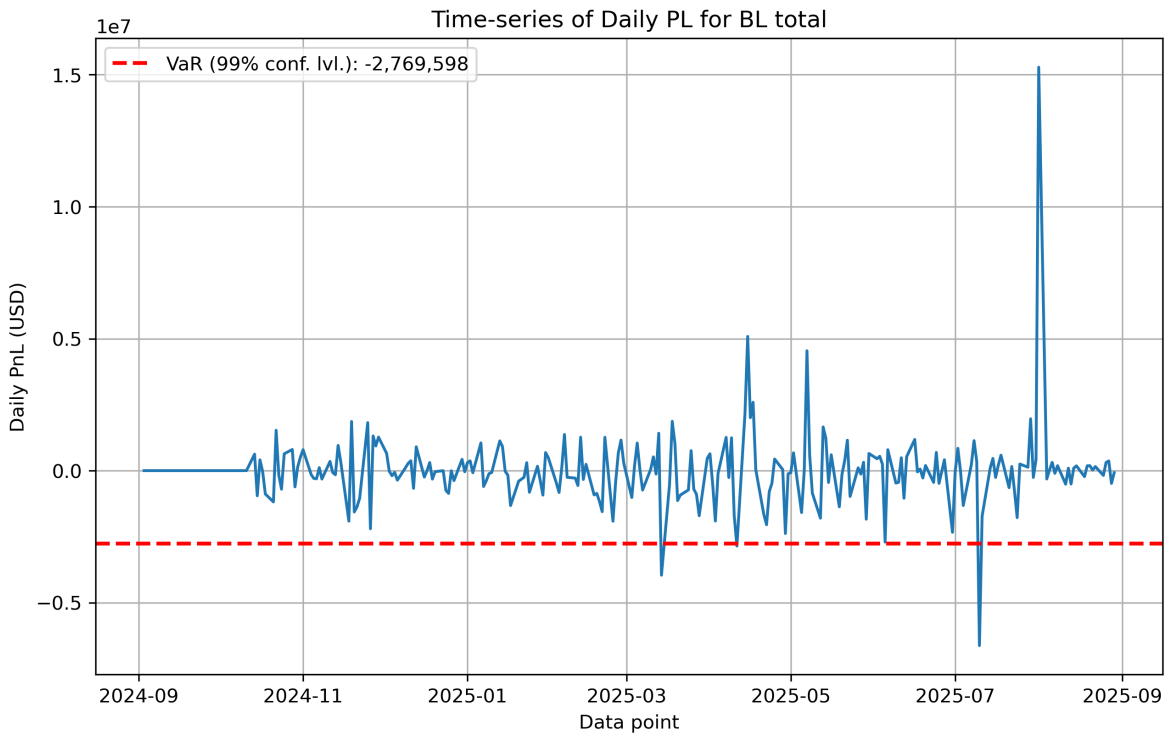

Figure 17: Across all business lines PnL

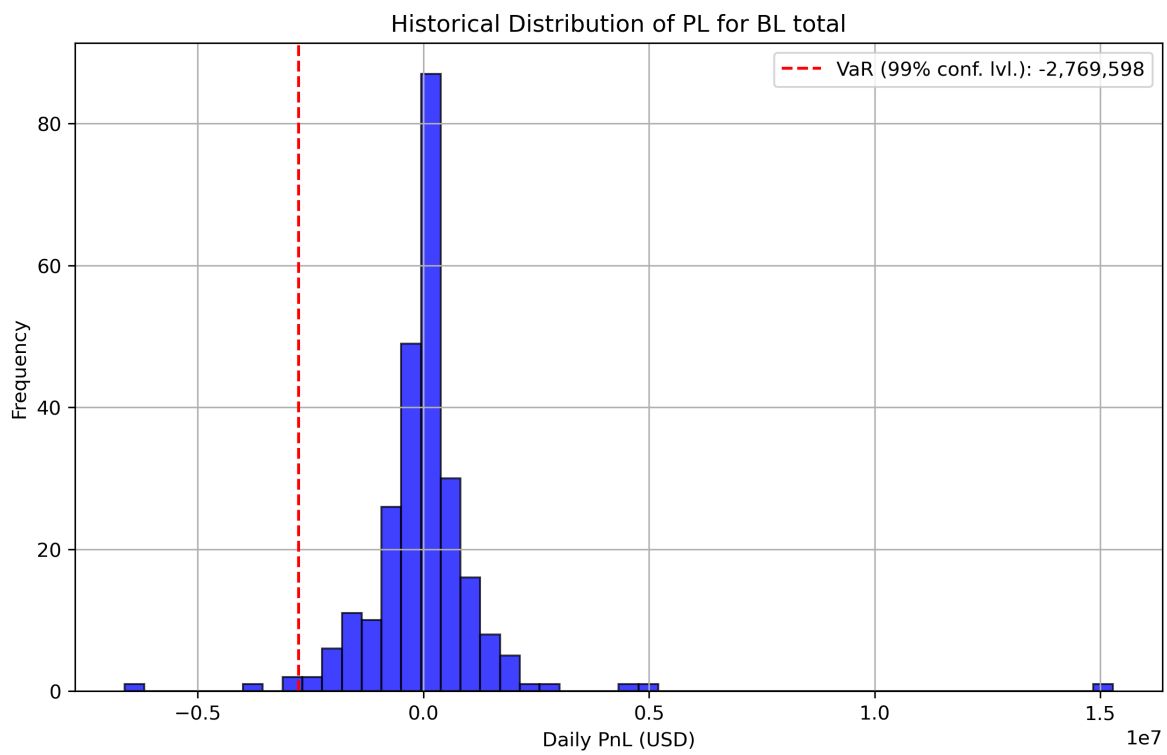

Figure 18: Across all business lines PnL Distribution