Simeng Wu

sw701@duke.edu | https://simengwu17.github.io/simeng-website/

EDUCATION

Duke University Durham, NC

Master of Science in Statistical Science Aug. 2025 - Expected May. 2027

Relevant Courses: Statistical Inference, Bayesian Analysis, Programming, Probabilistic Machine Learning

University College London London, UK

Bachelor of Science in Statistics, Economics, and Finance (First Class Honors, GPA: 3.94 / 4.0) Sep. 2022 - Jun. 2025

Relevant Courses: Calculus & Linear Algebra, Regression Analysis, Stochastic Methods, Medical Statistics

PROJECTS

Machine Learning-Based Option Volatility and Trading Strategy Development (Python) Feb. 2026 - Apr. 2026

- Built an SPY options volatility trading framework using GARCH models, Random Forest, LSTM, and neural networks, with features from historical volatility, trend indicators, option data, and VIX signals.
- Implemented and backtested a strangle options strategy on QuantConnect, taking long or short volatility positions based on forecasted volatility signals with defined entry, exit, and risk-management rules.
- Evaluated performance across in-sample, out-of-sample, and stress-test periods using return, Sharpe ratio, drawdown, and VaR; NN strategies achieved 16.8% and 20.9% OOS returns, while GARCH strategies kept drawdowns within 2% and 6%.

Momentum-Based Algorithmic Trading Strategy (Python) Mar. 2026 - Mar. 2026

- Developed a systematic momentum-rotation trading strategy in QuantConnect over selected S&P 500 stocks, using SMA, RSI, Bollinger Bands, and 20-day momentum signals to identify long opportunities.
- Conducted universe selection and exploratory analysis by screening stocks on liquidity, sector diversification, 60-day momentum, volatility, and maximum drawdown, selecting PLTR, GEV, and TSLA as the final candidate universe.
- Designed weekly rebalancing and daily risk-management rules, including SPY 200-day market regime filter, fixed stop-loss, trailing stop, trend deterioration exits, and RSI-based exits.
- Split the backtest into training, validation, and out-of-sample periods to reduce look-ahead bias; achieved an out-of-sample Sharpe ratio of 2.538 with net profit of $2.77M on $10M initial capital.

Travelers: Insurance Subrogation Prediction (Python) Oct. 2025 - Jan. 2026

- Built supervised machine learning models to predict 12,000 insurance subrogation outcomes using 30+ variables, including claim characteristics, loss severity indicators, and recovery-related attributes; ranked Top 9 on the public leaderboard.
- Performed end-to-end data preprocessing and feature engineering, including categorical encoding, interaction terms, and distributional transformations, which improved cross-validated AUC by 0.03.
- Developed and compared multiple tree-based models, including XGBoost, LightGBM, CatBoost, and TabM, and optimized hyperparameters through cross-validation and early stopping to improve robustness and reduce overfitting.
- Built an ensemble framework and automated the training and evaluation pipeline, improving out-of-sample consistency, reducing validation loss variance by 18%, and ensuring reproducible experimentation.
- Applied SHAP for model interpretability to identify key drivers of prediction outcomes and support model transparency.

Web-Scraped Recipe Dataset Construction and Interactive Analysis (R) Nov. 2025 - Dec. 2025

- Built an automated web scraping pipeline to collect and structure recipe data from the official game Wiki (Stardew Valley), covering recipe structure, ingredient composition, and numerical buff effects.
- Designed robust parsing and data validation workflows to handle heterogeneous HTML structures and inconsistent tag nesting, improving data extraction accuracy, reproducibility, and overall dataset quality.
- Developed core analysis logic using set operations and conditional filtering to support recipe search, craftable recipe identification, and minimal-additional-ingredient recommendations.
- Built an interactive R Shiny application that enabled flexible user exploration of the dataset and supported high-usability recipe analysis workflows.

Time Series and Copula-Based Monte Carlo VaR Modeling for a Two-Index Portfolio (Python) Mar. 2025 - Apr. 2025

- Built an equally weighted portfolio of the S&P 500 and SSE using weekly price data from 2000 to 2023.
- Transformed price series into weekly log-returns to stabilize variance and support ARMA-GARCH modeling assumptions.
- Modeled marginal return dynamics using R with ARMA-GARCH, selecting skew-t (S&P 500) and t (SSE) innovations via AIC/BIC and validating with PIT, KS/AD, and Ljung–Box tests.
- Captured cross-market tail dependence using a Clayton copula and generated 1,800 one-week Monte Carlo scenarios.
- Estimated 1-week portfolio VaR from simulated return distributions, yielding 99% VaR of -3.33% and 95% VaR of -2.13%.
- Evaluated VaR model reliability through back-testing considerations, focusing on tail loss coverage and model limitations.

Theme Park Simulation (Python) Jan. 2025 - Feb. 2025

- Developed an event-driven simulation framework in Python to optimize theme park operations, with the goal of reducing visitor wait times and improving operational efficiency and park throughput.
- Modeled the full visitor-facility flow, including arrivals, queueing, ride/service experiences, and facility transitions using a priority queue, exponential service times, and a stochastic transition matrix.
- Used pandas to analyze simulation outputs under varying arrival intensities, evaluating ride utilization, average waiting time, service efficiency, and other operational performance metrics.
- Incorporated customer preferences and ride constraints into the simulation, reducing 95th-percentile waiting time by 14% under peak demand.

Causal Analysis of Transportation Mode and In-Hospital Mortality in Trauma (Stata) Oct. 2024 - Dec. 2024

- Conducted an observational data analysis examining the association between helicopter versus ground ambulance transport and in-hospital mortality among trauma patients.
- Evaluated potential confounding by injury severity and age using stratified risk comparisons and crude versus adjusted effect estimates.
- Fit logistic regression models to assess how adjustment for measured severity altered the estimated association between transportation mode and mortality.
- Interpreted results with attention to confounding, covariate balance, and the limits of causal interpretation in non-randomized studies.

Analysis of a Randomized Trial of Video Consultation in Specialist Referral Pathways (Stata) Oct. 2024 - Dec. 2024

- Analyzed data from a multicenter randomized trial evaluating whether video consultation affected the probability of a subsequent follow-up appointment.
- Prepared trial-style summaries, including a CONSORT participant flow diagram and a baseline characteristics table, to describe enrollment, randomization, and follow-up.
- Estimated treatment effects using intention-to-treat (ITT) and covariate-adjusted logistic regression analyses, reporting effect measures on both absolute and relative scales.
- Investigated site-specific treatment-effect heterogeneity through interaction terms and considered the implications of missing data and non-compliance for interpretation.

Predict Hemoglobin levels with women's health data - Anemia (R) Mar. 2024 - Apr. 2024

- Conducted EDA on hemoglobin levels and demographic/environmental predictors, identifying non-linearity, multicollinearity, and potential interaction effects.
- Applied feature transformation, hierarchical clustering, PCA, and stepwise selection in R to improve model specification and reduce dimensionality.
- Built and diagnosed linear regression and Gaussian GLM models using residual analysis, Cook's distance, White's test, and Q-Q plots; refined the model with robust regression and WLS, achieving 70% adjusted R^2.

PROFESSIONAL EXPERIENCE

Da Hua International Management Consulting Co., Ltd. Shanghai, CN

Consulting Assistant Jun. 2024 - Sept. 2024

- Integrated and validated financial data across balance sheets, income statements, and cash flow statements; reconciled vouchers, contracts, and invoices to ensure accuracy and traceability.
- Supported financial due diligence for nearly 10 companies by reconciling accounts, analyzing three-year transaction trends, and preparing standardized reports and translated liquidation materials.
- Conducted EV charging infrastructure market analysis by scraping EV ownership and charging station data, assessing supply-demand gaps, identifying underserved cities for expansion opportunities, and developing PowerPoint presentations to communicate findings and recommendations to stakeholders.

Deyu Family Office Shanghai, CN

Investment Research Assistant Aug. 2023 - Sept. 2023

- Conducted wealth management industry research using SWOT analysis to assess market structure, competitive dynamics, and threats from traditional banks, helping identify core advantages and strategic growth priorities.
- Processed and analyzed financial and operating data using Excel and Python for data cleaning, metric calculation, trend visualization, and benchmarking against industry peers on growth, profitability, and cash flow performance.
- Supported valuation modeling by synthesizing multi-dimensional analysis, assessing competitive positioning and financial health, and providing data-driven insights for business decisions and strategic planning.

SKILLS

Analytical Skills	A/B Testing, Confounding Analysis, Optimization, Monte Carlo Simulation, Machine Learning
Technical Skills	Python (NumPy, pandas, scikit-learn, Matplotlib), R, SQL, Stata, SAS, LaTeX, Excel, PowerPoint