

Yin Zhang

(509) 339-3262
yin.zhang@wsu.edu

Data Scientist / Economist / PhD Student

GitHub: YinZhang0810
LinkedIn: yinzhang0810

EDUCATION

Ph.D. Econometrics and Quantitative Economic , <i>Washington State University</i>	Dec 2023[Expected]
M.S. Statistics , <i>Washington State University</i>	Oct 2023
M.S. Mathematics , <i>Washington State University</i>	May 2020
M.A. Financial Economics , <i>University of San Francisco</i>	May 2017
B.A. Finance (2nd Major: Management Operation) , <i>Washington State University</i>	May 2014

EXPERTISE

Time-Series Forecasting; Predictive Modeling; Causal Inference; Cross-sectional/Panel Data Analysis
Stochastic Risk Modeling; Machine Learning Classification; Data Driving Decision-Making;
Model Validation; Data Mining & Visualization; Anomaly Detection; Natural Language Processing;

SKILLS

Tools and Languages	Python, R, SQL, ArcGIS, SAS, Stata, \LaTeX , Markdown
Packages	Pandas, NumPy, NLTK, scikit-learn, TensorFlow, PyTorch, PyOD, StatsModels, Matplotlib, Git
Communication	English(Fluent), Chinese (Native)

TECHNICAL EXPERIENCE

The Impact of Online Investor Sentiment on Cryptocurrency Return [Python] **Working Paper**
An Application of NLP and ML Techniques on Financial Asset Return Prediction *Washington State University*

- Scraped 10 million+ cryptocurrency related tweets by Twitter API and corresponding financial data. Applied basic data pre-processing, including missing value check, outlier detect, stationary check, visualization, etc.
- Cleaned text data by tokenization and normalization using Natural Language Processing (NLP) toolkits.
- Classified sentiments and assigned polarity scores of sentiments for each tweet by both lexicon-based algorithm and supervised classifier (i.e. logistic regression)
- Examined causal relationship between sentiment polarity scores and cryptocurrency return by Granger-Causality test.
- Obtained significant predictive power of Twitter sentiment on cryptocurrency returns by adopt different time-series forecasting methods (i.e. VAR, XGBoost, RNN, Prophet) with and without sentiment polarity scores.
- Applied validation process at both sentiment analysis, causality analysis, and forecasting parts to ensure the accuracy of polarity scores and validity of findings.

Analysis the Impact of Adverse Weather Conditions on Flight Delay Problem [Python] **Working Paper**
Washington State University

- Analyzed a dataset contains 7.6 million flight departure and arrival information and corresponding airport weather information including rainfall, snowfall, and wind at different intensity.
- Performed exploratory data analysis (EDA) and conducted inferences at both daily and hourly level of collected data. Selected the thresholds of adverse weather condition based on existing literature or exploratory findings.
- Estimated the impacts using logistic regression and difference-in difference estimator.
- Conducted a series of robustness tests on estimated results to ensure the validity of the findings.
- Discussed the costs/economic losses due to flight delays from various perspectives (i.e. air passenger, freight transport, airline, and whole economy) based on empirical finding, which would provide insights for policymakers' decision-making process.

Incident Topic Similarity Matching and Its Application to Accounting Information Retrieval [Python] **February-May 2022**
Microsoft *Remote*

- Performed dictionary build, frequency count, and supervised classifier with incident data from Microsoft Enterprise Data Lake House.
- Applied NLTK POS tagger. Evaluated the improvement of classification accuracy by removing irrelevant words.
- Adopted Sentence-Bert for sentence classification. Adopted Facebook fast text for multi-intent classification.

Anomaly Detection and Future Usage Forecasting on Storage [Python] **February-May 2022**
Microsoft *Remote*

- Performed Exploratory Data Analysis (EDA) for daily time series data obtained from Microsoft Enterprise Data Lake House. Visualization were plotted using matplotlib and plotly.
- Detected anomalies using PyOD API (Median Absolute Deviation, K Nearest Neighbors, Local Outlier Factor, Isolation Forest)
- Per-processed data by replaced the anomalies and missing values.
- Predicted future storage capacity, read/write operation size using (1) ARIMA with per-processed data (2) Prophet with change points detected. Selected the best combination of a grid of parameters via Time-Series split cross-validation.

Yin Zhang

(509) 339-3262
yin.zhang@wsu.edu

Data Scientist / Economist / PhD Student

GitHub: YinZhang0810
LinkedIn: yinzhang0810

OTHER RESEARCH & COURSE PROJECTS

Credit Card Fraud Detection Using Supervised Classifiers [Python] *Machine Learning Course Project*

- Performed EDA on given dataset. Resampled the imbalance data. Split dataset as train and test data. Conducted the Native Bayes, Random Forest, Logistic Regression and Support Vector Machine (SVM) methods to classify the fraud. Fine-tuned their hyperparameters to get better results. Evaluated different models by several metrics.

Forecasting JPY/USD and GBP/USD Exchange Rate by ARIMA model [SAS and R] *Time Series Analysis Course Project*

- Utilized the classic ARIMA model by exchange rate data, including data pre-processing, model identification by ACF and PACF plots, parameter estimation, model validation, forecasting, and model selection.

Pricing European Options Under Heston Model [R] *Capstone Project of M.S. Mathematics*

- Estimated parameters by Maximum Likelihood, calculated option price, checked validation, and discussed properties of the Heston model. Increased the forecasting accuracy by 13%, compared with the classical Black-Scholes option pricing model.

Project of PNW Container Optimization Model *Research Volunteer*

- Collected, cleaned, and prepped data using Stata, SQL, and Excel to help Professor build container rental and transportation optimization model. Contributed towards the development of optimization model and conducted econometric analysis on operational costs reduction problem.

Applied Bayesian Approach to Predict Carbonation Depth of Concrete [R] *Bayesian Analysis Course Project*

- Collaborative project with Civil Engineering researcher. Computed the likelihood function and posterior distribution, and predicted the carbonation depth of concrete. Evaluated the model by comparing the results to actual experimental observation.

PROFESSIONAL EXPERIENCE

Solo Instructor

August 2020-Present

Washington State University

- EconS 326: Aspects of Sustainable Development (online)
- EconS 351: Introduction to Food and Agricultural Markets (online)
- EconS 321: Economics of Sports in America (online)

Fall 2020/2021/2022

Spring 2021/2022/2023

Summer 2021

Graduate Teaching Assistant

August 2017-May 2020

(leading review session, grading assignments, holding office hours)

Washington State University

- EconS 101/102: Undergrad Micro & Macro
- EconS 450: Advanced Farm and Ranch Management
- Stat 443: Applied Probability
- Stat 412: Statistical Methods in Research
- Math 106: College Algebra
- Math 171(Lab) Calculus 1
- Tutor at Math Learning Center, WSU

Fall 2019/Spring 2020

Spring 2019

Fall 2018

Fall 2018

Spring/Summer 2018

Fall 2017

Graduate Teaching Assistant

January 2016-May 2017

(leading review session, grading assignments, holding office hours)

University of San Francisco

- Econ 655: Options & Futures
- Econ 691: Special Topic: Financial Derivatives
- Econ 101/102 Undergrad Micro Macro

Spring 2017

Spring 2017

Spring/Fall 2016

ACTIVITIES

Python Working Group

Active Member / Presenter

Washington State University

January 2020-Present

Pacific Northwest National Laboratory (PNNL)-WSU Join Seminar

Presenter

Washington State University

October 2022

School of Economic Science (SES) Student Seminar

Presenter

Washington State University

May 2022

American Economic Association

Member

American Statistical Association

Member