Robertson Wang

Education

2008–2012 **BSc, Economics**, *New York University*.
Magna Cum Laude

2015–2016 **MSc, Economics**, *Universitat Pompeu Fabra*.
Fundación Ramón Arece Scholarship

2016–2018 **MSc, Data Science** , *George Washington University*.

2017 **Data Science Certificate**, *Georgetown University*.

Summary

Data science
My primary programming languages are Python and R. My main interests are in applications of machine learning and natural language processing. I am currently working on academic research using sentiment analysis and machine learning. I've also been involved in machine learning initiatives applying computation linguistics, neural networks, and NLP to banking regulation and identifying emerging financial risk. I am well versed in sklearn, TensorFlow, SciPy, and nltk.

Day to day I use a variety of statistical modelling techniques and machine learning algorithms. These include least-squares regression, logistic regressions, neural networks, convolutional neural networks, support vector machine based classifiers, various bag of word models, Latent Dirichlet Allocations, word vectors, and XGBoost (Regression Tree Classifiers).

Experience

2016–2017 **Research Assistant, Risk Analysis**, *Federal Reserve Board*, Washington DC.
Python & R Developer

- Co-Authored an academic paper using sentiment analysis to capture settings in which management communications have greater credibility
- Wrote web-scrappers in Python using BeautifulSoup, lxml, and selenium to acquire corpus of financial documents.
- Managed large unstructured datasets (10 tb+) on Hadoop, wrote BASH scripts and cron jobs to streamline the data pipeline.
- Developed code in Apache Spark to quickly process large amounts of natural language data, such as text from 10-K and earnings announcements.
- Put into production NLP-based measures, built using Latent Dirichlet Allocation and word vector skip-gram models, of emerging financial risk to help policy makers identify areas in need of stricter oversight
- Produced web-scrapping software in order to perform on-the-fly NLP analysis and classification of press releases with the goal of monitoring emerging areas of risk
- Performed entity resolution and data cleaning on time series CDS positions data (100+gb/day) in order to produce quarterly risk profiles on FRB regulated banks
- Wrote Python and R code to work with the SLURM cluster and Apache Spark in order to parallelize production work, reducing overall costs and lag time in producing results

135 Adams Street NW – Washington, DC 20001
1 (908) 745 8075 • *robertsonwang@gmail.com*
www.linkedin.com/in/robertsonwang/ • *robertsonwang.github.io*

2012–2015 **Financial Associate, Valuations & Modeling**, *Morgan Stanley*, New York.

R Developer & Master of Excel Hell

- Crafted various regression models for asset pricing. Types of regression models that we used includes LASSO, Ridge, SVAR, and least-squares.
- Created Bayesian forecasting models to price mortgage and credit derivatives
- Wrote code for a high performance cluster in order to price securities using Monte-Carlo and other simulation techniques. Used SLURM and Apache Spark.
- Created quantitative models to value securitized products (MBS, ABS, MBS Derivatives)
- Led methodology development on a option pricing model, based on Black-Scholes, to price a bespoke REIT IPO.
- Designed databases, stored procedures, reports, and data input interfaces using SQL Server 2012, 2016 and PostgreSQL.
- Automated reporting tasks using Python, VBA, and SQL.
- Wrote research reports on macroeconomic risk, local job market/housing market conditions, and real estate specific background.
- Worked with external auditors and firm counter parties to communicate and defend parameter estimations and model assumptions.
- Managed and provided mentorship to a team of financial analysts and summer interns

Master thesis

Title *Rational Bubbles And the Financial Accelerator*

Supervisor Alberto Martin

A game theoretic approach to financial bubbles, wherein we model bubbles as a Pareto optimal outcome to a repeated coordination game.

Capstone Project

Title *A NLP Restaurant Recommendation System Based On Yelp Reviews*

We used a combination of LDA and tf-idf features trained on user reviews to recommend new restaurants to users. For each user, we test three different classification algorithms and select the best performing machine learning algorithm. We used XGBoost, Random Forest, and linear SVM.

Technical Skills

Statistical Packages	Matlab, R, Stata	*Data Stores*	SQL, MongoDB, SAS
Programming	Python, C++, Java	*Big Data*	Apache Spark, Hadoop, Slurm
Markup	LaTeX, XML, MarkDown	*Office*	Excel, PowerPoint, Word

Relevant Coursework

- Machine Learning
- Probability Theory
- Time Series Analysis
- Data Visualization
- Regression Models & Panel Data
- State Space & Bayesian Models

Interests

Rock Climbing, Modern Literature, and Linguistics

135 Adams Street NW – Washington, DC 20001

1 (908) 745 8075 • *robertsonwang@gmail.com*

www.linkedin.com/in/robertsonwang/ • *robertsonwang.github.io*