

Experience

Harvard University, Growth Lab

Feb. 2016 - present

Research Fellow, project lead (2019 - present)

- Manage team of research analysts / assistants [3FTE]
- Lead researcher, under review at *Science*: Extracting insights from 324 historical books on patents (1790 - 1954)
 - Developed denoising CNN to remove noise from scans
 - Developed LSTM neural net on historical fonts for optical character recognition (OCR)
 - Re-trained BERT model for named entity recognition (NLP)
 - Created parallel-computing libraries to scale models to ~500k pages and link output to ~10B Census records
 - Adopted university-wide by Harvard, integrated into supercomputing's SLURM scheduler (CentOS)
- Led project for World Bank on extracting investment insights from new dataset on ~120 million firms
 - Fine-tuned multimodal Transformer model to extract input-output data from ~50K company invoices
 - Developed and implemented matching algorithm on World Bank's servers to link firms to ~100 million patents
 - Led to follow-up work for a major aerospace company and a major beverage company
- Created data science committee for:
 - Moving lab's workflows, data (~300TB) and 76 user envs. across UNIX envs. to supercomputing cluster
 - Managing lab's requirements with Amazon AWS, Google BigQuery, Dun Bradstreet (~\$75K annually)
 - Weekly meetings to discuss / implement scientific advances
- Contributor on GitHub to open-source data science libraries
 - Polars (Python / Rust, Pandas alternative): created Jupyter implementation for Emacs
 - EconGeo (R, calculates eigenvalue centrality algorithms): optimized and refactored for CRAN submission
- Developed and taught Python / R sessions (10 annually) to MPA students with backgrounds in linear algebra
 - Topics: Eigenvector interpretations of economic complexity, machine learning (OCR/NLP), parallel computing

Postdoctoral Fellow (2016 - 2019)

- Econometric research on machine learning, network analysis, causal inference. 700+ citations Google Scholar
 - Implemented forecasting models on-site at statistical offices in Mexico, Panama and Saudi Arabia
 - With digital design team, created dashboard for Mexican government on labor mobility predictions
- Teaching assistant in course 'Executive Education: Leading Economic Growth'
 - Guided senior government officials and executives on Economic Complexity analysis

Education

Ph.D. in Computational Economics, Utrecht University	2016
• Longitudinal research on Statistics Sweden's supercomputing cluster on ~20 billion administrative records	
• Teaching: Econometrics (Master), Statistics, Organizational Networks (coordinator), Applied Economics	
M.Sc. Local Economic Development, London School of Economics and Political Science (with merit)	2010
M.Sc. Human Geography and Planning, Utrecht University (cum laude)	2010
B.Sc. Human Geography and Planning, Utrecht University (cum laude)	2008
• Teaching assistant in statistics, research assistant in interviewing 65 aerospace companies	

Awards

Regional Science Association International Taught Postgraduate Student Award	2011
• Best Master thesis in UK/Ireland ("Using GMM estimators and dynamic panel regressions for causal inference")	
Huygens Talent Scholarship (by the Netherlands' government to study at London School of Economics)	2009

Skills

Programming languages: Python, SQL, R, STATA, Lisp (Emacs), Lua (Neovim), Rust (basic)

Tools: Analysis: Numpy, Pandas, Polars. Visualization: Matplotlib, Plotly, ggplot2. Machine learning: PyTorch, SciPy stack, Hugging Face, LayoutVM2, NLTK, Tesseract, spaCy. Data: PostgreSQL, DuckDB. Workflow: UNIX tools, git

Modeling techniques used: Deep learning, Transformers, multi-modal neural networks (OCR / NLP), cross-validation, boosting (XGBoost), exponential random graph models, regression, instrumental variable analysis, robust statistics, A-B testing / experimental designs / RCTs, GMM estimators, robust statistics, hazard models