

Li Shen, PhD

Worcester, MA | lshen@clarku.edu | (716) 817-3989 | [linkedin.com/in/lily-shen](https://www.linkedin.com/in/lily-shen)

SUMMARY

Ph.D.-trained applied economist who translates pricing, market, and regional data into decision models. Brings complementary strengths in pricing-demand modeling and decision-support tools, structural spatial and trade research with calibrated counterfactuals, and panel-data analysis. Builds auditable workflows from multi-million-row data using explicit assumptions, validation checks, and reproducible Python, SQL, Stata, MATLAB, R, and SAS code.

EDUCATION

Clark University <i>Ph.D. in Economics</i>	Worcester, MA 2026
The State University of New York at Buffalo <i>M.A. in Economics</i>	Buffalo, NY May 2019
Valparaiso University <i>M.S. in International Economics and Finance</i>	Valparaiso, IN May 2018
Chongqing University <i>B.Mgmt. in Accounting</i>	Chongqing, China June 2014

TECHNICAL SKILLS

- **Applied econometrics:** Panel-data methods, difference-in-differences, shift-share IV, staggered adoption designs, event studies, gravity models, and time-series forecasting (ARIMA).
- **Structural and decision modeling:** Quantitative spatial equilibrium, quantitative trade, multi-region input-output networks, calibration, labor mobility, counterfactual simulation, constrained optimization, and sensitivity checks.
- **Programming:** Python, SQL, Stata, MATLAB, R, SAS.
- **Data and delivery:** pandas/Polars, GeoPandas, ArcGIS, Streamlit, Git/GitHub, pytest, ggplot2, LaTeX; experience presenting technical results to academic and nontechnical audiences.

RESEARCH AND ANALYTICAL EXPERIENCE

Pricing and Promotion Decision Engine — Demand Estimation, Optimization, and Deployment 2025–Present

Live app: pricing-promotion-decision-engine.streamlit.app Code: github.com/Li-Shen-Clark/pricing-promotion-decision-engine

- Built an end-to-end pricing decision-support application on 6.6M rows of Dominick’s Finer Foods scanner data; cleaned to a 4.65M-row panel covering 486 UPCs, 93 stores, and 366 weeks.
- Estimated a log-log fixed-effects demand model with Duan smearing retransformation to forecast demand responses to price and promotion changes.
- Implemented counterfactual price and promotion simulations plus constrained profit optimization under cost, inventory, and margin constraints.
- Designed A/B validation plans with power analysis for raise-and-test pricing candidates; deployed as a Streamlit app with modular Python code, reproducible requirements, GitHub auto-deployment, and pytest tests.

Vegetable Price Intelligence — Forecasting, Risk, and Model Governance 2026–Present

Portfolio case study: [Case Study](#) Code: [GitHub](#)

- Built an economics-informed price intelligence system from 8.68M historical wholesale-market records across 117 cities and 30 vegetables, producing 7.36M auditable city-day facts with versioned mapping, quality flags, and coverage tiers.
- Developed rolling-origin 7/14/28-day forecasts against persistence and seasonal baselines; on a sealed 16,217-task final test, reduced overall WAPE from 18.72% to 17.54% (6.30%) and routed 10 weak product-horizon groups to explicit baseline fallback.
- Built an independent 14-day price-spike classifier with a validation-frozen threshold, reaching PR-AUC 0.2204 versus 0.0775 for the strongest simple baseline, 37.69% recall at 9.99% FPR, and 7.18 days of average historical lead time for human review.
- Evaluated 2,736 predeclared city lead-lag candidates using leave-one-out common factors, product-level BH-FDR, split stability, 250 moving-block bootstrap replications, and sealed final testing; withheld the directional network after only 4 of 15 frozen edges improved out of sample.

Structural Spatial Economics and Quantitative Trade — Trade Shock Propagation 2023–Present

- Co-developed a multi-sector structural spatial-equilibrium model (31 provinces, 3 sectors) with inter-provincial input-output networks, non-tradable production, and labor mobility, calibrated to China’s 2017 MRIO table; contributed model development, structural parameter recovery, calibration-data work, and counterfactual analysis.
- Established local trade multipliers as equilibrium objects shaped by domestic production structure rather than universal parameters, identifying the own-trade share—not raw export exposure—as the structural organizing statistic for multiplier heterogeneity.
- Extended the quantitative spatial economics framework from national welfare analysis to sub-national shock amplification; counterfactual simulations yield province-level GDP multipliers ranging from 0.30 to 1.02, showing that high local absorption amplifies external shocks through non-tradable service demand—reversing the conventional exposure-based intuition.

High-Frequency Price Analytics — Infrastructure and Market Integration

2021–Present

- Constructed the largest daily-frequency panel of Chinese vegetable wholesale prices to date: 3.99 million city-pair-month observations covering ten commodities across 117 prefecture-level cities (2014–2022), merged with AutoNavi smartphone-based intercity passenger mobility indices.
- Leveraged two institutional features—the Green Passage highway-toll exemption (fixing variable freight costs for fresh vegetables) and prohibitive HSR freight tariffs for bulk produce—to study a setting in which differences in HSR connectivity reflect variation in passenger rather than freight costs.
- Documented that HSR-connected city pairs exhibit 2–5% lower intercity price dispersion; passenger-mobility controls absorb the direct HSR coefficient, and cross-product heterogeneity (larger effects for bulky, low-value commodities) argues against a freight substitution channel.

China Population Density Atlas — Interactive Geospatial Product

2022–Present

Live app: pop-density-spatial.streamlit.app

- Processed 19 years of 5 km-resolution population-density rasters (2002–2020) covering China into a harmonized annual panel for grid-cell and province-level analysis.
- Quantified increasing spatial agglomeration and the expansion of dense urban corridors; found little evidence of province-level convergence.
- Deployed an interactive Streamlit application in 2025 for map-based exploration of density, agglomeration, and urban-corridor metrics; live with auto-deploy from GitHub.

Demand Forecasting and Revenue Modeling

2017–2019

- Built seasonal ARIMA models in SAS to forecast international passenger demand; estimated income elasticities for capacity planning.
- Modeled determinants of Intel Corp.’s net revenue using regression analysis; coauthored a publication in *Engineering and Technology Quarterly Reviews*.

KAGGLE COMPETITIONS — APPLIED MACHINE LEARNING

Home Credit — Credit Risk Model Stability (Silver Medal)

Feb 2024–May 2024

Role: Statistical modeling & co-developer Stack: Python, Polars, Pandas, scikit-learn, LightGBM, CatBoost

- Co-developed a high-performance ETL pipeline in Polars using a lazy-execution strategy to process multi-depth relational data; designed aggregation features (max, last, mean) capturing borrowers’ historical behavioral trends, adopted across the team codebase.
- Guided a correlation-based feature-grouping procedure (Pearson threshold > 0.8) to address multicollinearity and filter uninformative variables; enforced memory optimization via type downcasting, reducing memory usage by over 50%.
- Optimized directly for the competition’s headline Gini Stability metric, ensuring that model performance did not degrade over time.

ICR — Identifying Age-Related Conditions (Bronze Medal)

May 2023–Aug 2023

Role: Machine-learning researcher & co-developer Stack: Python, TabPFN, XGBoost, imbalanced-learn

- Designed a stratified resampling strategy using auxiliary “Greek” metadata: applied RandomOverSampler on latent subgroups (Alpha variable) rather than simple oversampling, preserving the underlying population structure under severe class imbalance.
- Co-developed a voting ensemble combining parametric (XGBoost) and non-parametric (TabPFN) models under a custom Balanced Log Loss metric; fine-tuned decision-boundary calibration to minimize false negatives—critical for medical-diagnosis tasks.

ADDITIONAL EXPERIENCE AND RECOGNITION

- **Teaching Assistant, Clark University** (2020–2025): led computational sessions on Monte Carlo simulation, stochastic processes, and portfolio analysis for graduate finance courses; supported undergraduate economics instruction.
- **Honors:** Sheftel Award for Research Excellence (2025), Veendorp Best Field Paper Award (2022), Doctoral Fellowship (2019–2026).