Stress Testing Machine Learning Population Change Projections with ACS MOEs

By Collin McCarter

Opinions and ideas are my own disclaimer ... Feel free to reach out on LinkedIn for any data digression!

Southern Demographic Association Annual Meeting, October 20th, 2023

https://www.linkedin.com/in/collin-mccarter/ ... https://github.com/collinmccarter-fl/stresstest_ACS_MOEs/

2022 - TODAY ... Data Science, Northwestern Mutual
2020 - 2022   ... Technology & Product Management, Trellance (Credit Unions)
2016 - 2020   ... Demographer, Nielsen (TV Ratings)
2014 - 2016   ... PK-12 School Enrollment Analyst, State of Florida

Background

Life as a data scientist ...

- Organizations **link ACS data with a "client dataset"** by small-area geographies
- Organizations are always searching for **"new data sources"** to differentiate the organization
- Discovering "new data sources" comes with a **learning curve**
- ACS MOE's can be a **time consuming** learning curve for big data practitioners, especially those without a survey background
- **Stress testing examples** may help to lessen the learning curve, which is possible with MOE-adjusted ACS variables

Life as an applied demographer ...

- Open source tools are **efficient, accessible, and reproducible** but are infrequently used in applied demography
- PIP packages are **maturing** in their documentation, peer-review credibility, and flexible options.
- Machine Learning practitioners are using these tools and data **without domain knowledge**
- Applied Demographers **have the domain knowledge** to maximize value from machine learning tools used in demographic analysis

# What I did ... stress tested ACS1yr CA County projections

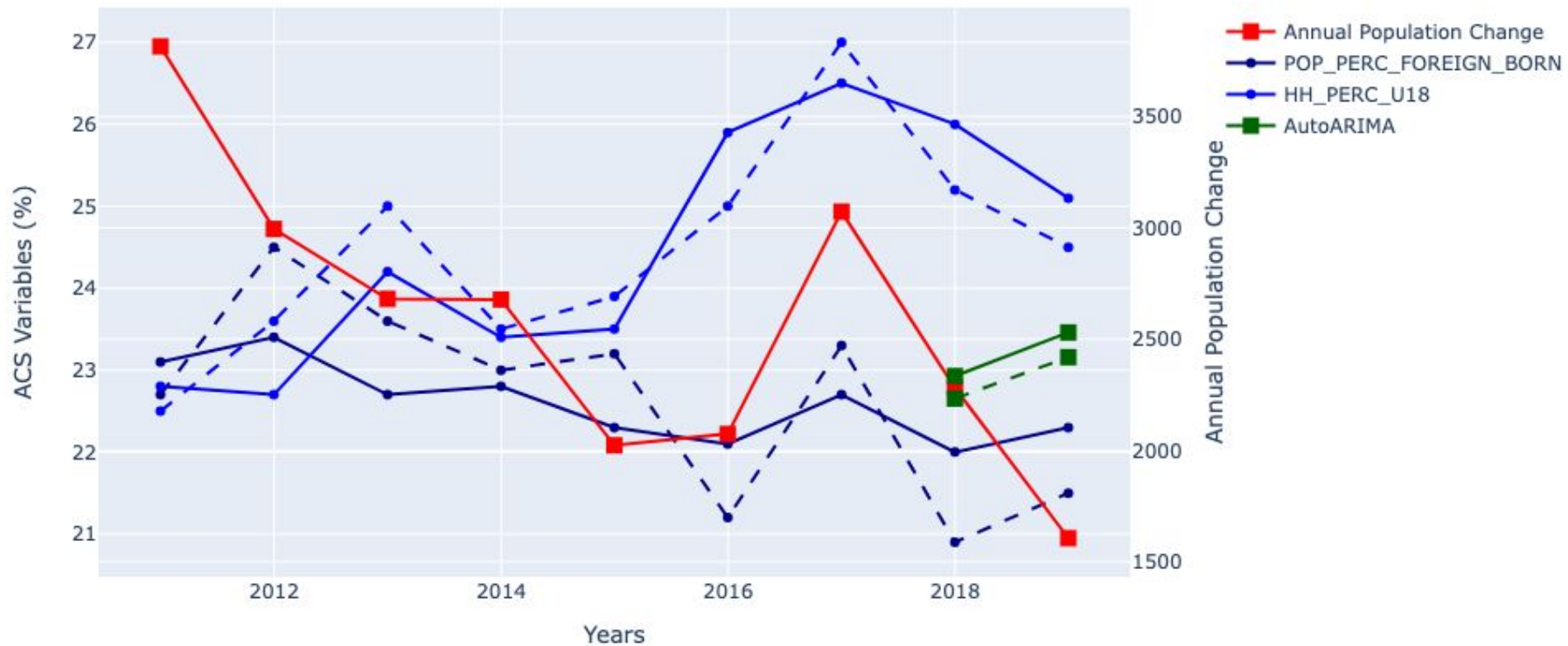Python Jupyter Notebook completes the below in **< 3min**

- *1 sec* Gather packages of functions (aka procedures) from PIP
- *9 sec* Pull select api.census.gov data
  - ACS 1yr Data Profile
  - PEP Vintage 2010
  - PEP Vintage 2020
  - PEP Vintage 2022
- *1 sec* Transform and Merge US Census data into a training dataset for machine learning packages
- *150 sec* Train (fit), Predict, and Evaluate different projects based on below adjustments:
  - ACS variables used as predictive features vs univariate timeseries
  - Adjusted ACS variables using MOEs vs ACS variable point estimates
  - Continuous Regression metrics (MAE, MSE, MAPE, r^2)
  - Models available on PIP (structureboost, statsforecast, catboost)
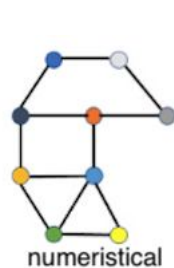
Los Angeles County 2yr Population Change Projections, Adjusting ACS Vars with MOEs (dashed)
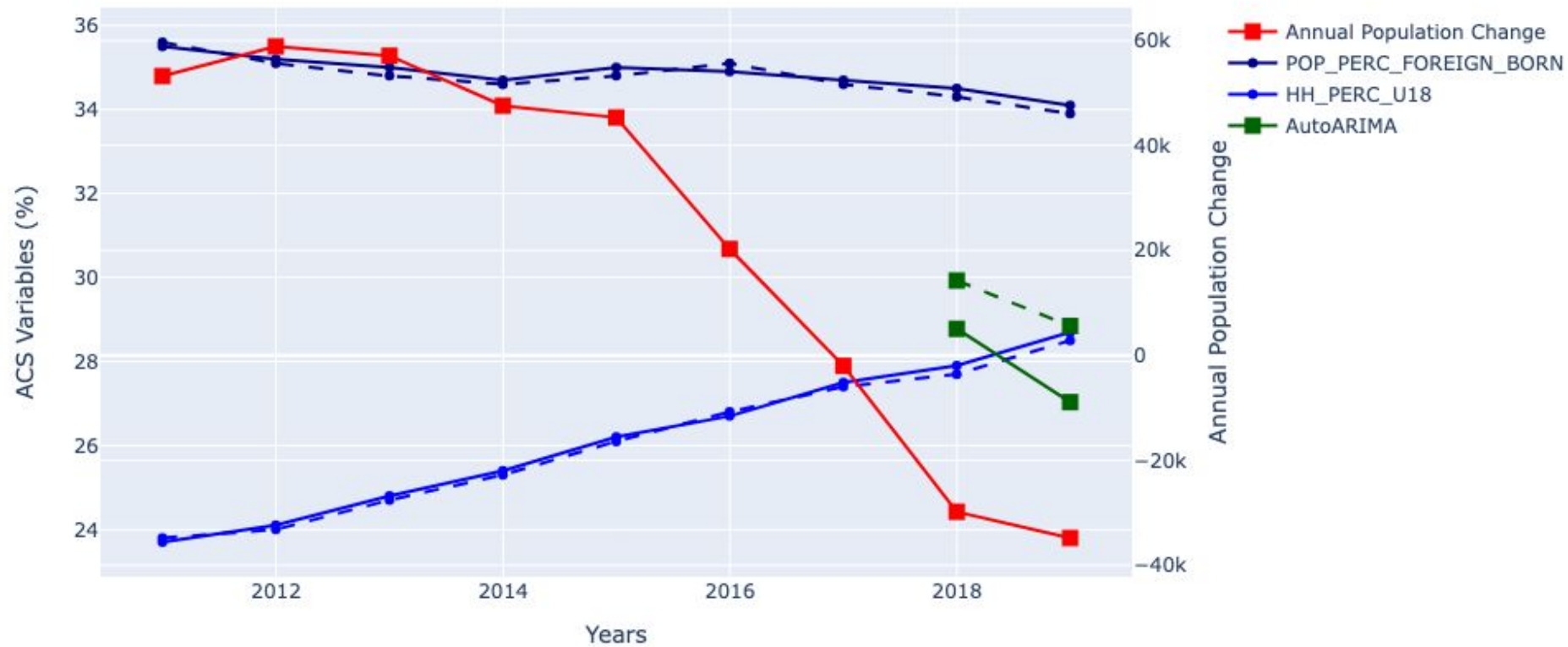
Tulare County 2yr Population Change Projections, Adjusting ACS Vars with MOEs (dashed)

Thank You!

Collin McCarter