Discussion: Expanding the Frontier of Economic Statistics Using Big Data

A Case Study of Regional Employment

Discussant: Max Ghenis

PolicyEngine

Society of Government Economists
April 4, 2025

POLICY ENGINE

Paper by: Abe Dunn (BEA), Eric English (Census), Kyle Hood (BEA), Lowell Mason (BLS), Brian Quistorff (BEA)

POLICY ENGINE

- **Objective:** Evaluate using third-party payroll data ("Big Data") to enhance official regional employment statistics (State/County-by-Industry).

- **Core Problem:** Leveraging new data sources amidst declining survey response rates, balancing timeliness/granularity vs. accuracy.

- **Framework:**
  - Production Possibility Frontier (Accuracy vs. Granularity/Timeliness).
  - Using existing official statistics' revision errors (CES vs. QCEW) as a benchmark/tolerance level.

- **Methodology:** Linear models predicting QCEW growth using CES & payroll data features; out-of-sample validation (Rolling & K-Fold).

- **Key Findings:**
  - Payroll data improves state-level accuracy modestly (11-19% MAE reduction in areas).
  - New county-level estimates generated meet extrapolated accuracy standards.
  - Demonstrates value for timeliness/targeting in COVID application.

- **Timely & Important:** Addresses a critical challenge for statistical agencies (integrating alternative data).
- **Innovative Framework:**
  - Systematic use of existing revision errors ("error tolerance") as a benchmark is clever and practical.
  - Novel statistical test for evaluating new granular series against the existing frontier.
- **Methodological Rigor:**
  - Careful construction of payroll series (continuing sample).
  - Use of appropriate benchmarks (QCEW).
  - Robust cross-validation techniques (Rolling, K-Fold).
- **Practical Relevance:** COVID-19 application demonstrates potential real-world value for policy.

POLICY ENGINE

- **Successful Frame:** Paper effectively uses the lens of "how can we best predict QCEW?" to integrate payroll data with CES.
- **Machine Learning Explored:**
  - Paper briefly compares linear models to Random Forest (RF).
  - Finds mixed results for RF: Good performance in cross-fold validation, but poor in rolling validation (especially during COVID).
  - Hypothesis: RF struggles with extrapolation for unprecedented events compared to linear models in this test.
- **Question:** Is this performance inherent to RF, or partly due to the limited features used in the test?

- **ML thrives on predictors:** Could ML models (RF, Gradient Boosting, etc.) perform better with a richer feature set?
- **Leveraging Payroll Data Further:**
    - Include features from *both* "Continuing Units" and "Full Data" samples to capture stable trends *and* entry/exit dynamics?
    - More detailed payroll characteristics (firm size mix in sample, data lags, etc.)?
- **Leveraging CES Data Further:**
    - More historical CES data? Sectoral trends?
- **Other Data Sources:**
    - Incorporate related signals? (E.g., relevant CPS employment measures, spatial/neighboring area trends). ML can often effectively sift through many predictors.

- **Beyond Point Estimates - Prediction Intervals:**
  - Suggestion: Use Quantile Regression Forests (QRF) to predict the *distribution* of potential QCEW outcomes.
  - QRF PIs capture uncertainty relative to the QCEW benchmark (sampling, non-sampling, model error combined).
  - Complements official CES CIs (which focus on initial sampling error).
- **Beyond MAE - Heterogeneity & Constraints:**
  - Paper focuses on average accuracy (MAE).
  - Are there institutional needs to consider other metrics? E.g., minimizing instances where the combined model is *worse* than raw CES, even if MAE improves? (Loss aversion to increased error in specific cases).

- **Viewing CES as QCEW Predictor:** If a primary goal of timely CES estimates is to predict the eventual QCEW benchmark...
- **...What does this imply for optimizing CES itself?**
  - **Survey Weights:** Could CES sample weights be optimized differently if maximizing predictive accuracy for QCEW (perhaps conditional on payroll data) is an explicit goal alongside representation?
  - **Survey Design/Content:** Could survey questions be added or modified to gather information known to improve QCEW prediction accuracy?
  - Galvanizes thinking about the entire data production pipeline through a predictive lens.

POLICY ENGINE

- **Valuable Contribution:** Excellent paper demonstrating a framework and application for integrating alternative data, successfully expanding the "frontier".
- **Key Theme:** Framing CES/Payroll as predictors of QCEW opens productive avenues for research and potentially operations.
- **Future Directions Highlighted:**
  - Deeper ML exploration (more features, incl. full continuing payroll).
  - Quantifying prediction uncertainty (QRF for PIs).
  - Considering broader evaluation metrics (beyond MAE).
  - Exploring implications of predictive goals for CES survey design/weighting.
- **Congratulations to the authors!**