Improving Macroeconomic Model Validity and Forecasting Performance with Pooled Country Data using Structural, Reduced Form, and Neural Network Models

Cameron Fen and Samir Undavia

Abstract
We show that pooling countries’ macroeconomic data across a panel dimension produces a statistically significant improvement in the generalizability of structural, reduced form, and machine learning (ML) methods, producing state-of-the-art results. Using GDP forecasts evaluated on an out-of-sample test set, this procedure reduces RMSE anywhere from 12% to 24% depending on the type of model. Forecasting using non-US-pooled-data, we show that reduced-form and structural models are more policy-invariant and outperform a US-data-only baseline. Our deep learning approaches outperform all tested baseline economic models. Robustness checks indicate that our outperformance is reproducible, numerically stable, and generalizable across models.

Models Used
- **Var(4)** - A 4-period linear autoregressive model using GDP, consumption, and employment as inputs
- **AR(2)** - A 2-period linear autoregressive model using only GDP
- **Smets-Wouters DSGE** - Standard DSGE from the Smets-Wouters paper
- **Factor** - A linear model that takes in 248 data series and uses PCA to reduce dimensionality
- **Our Custom RNN** - A gated-recurrent-unit-based neural network with skip connections and dropout
- **AutoML** - An algorithm that chooses the best machine learning model from a collection of models

### Structural and Reduced Form Models

| Time (Q's Ahead) | 1Q | 2Q | 3Q | 4Q | 5Q |
|------------------|----|----|----|----|----|
| **VAR(4)**       |    |    |    |    |    |
| US Data          | 2.99 | 3.03 | 3.10 | 3.08 | 3.08 |
| World Data       | 2.37 | 2.52 | 2.56 | 2.63 | 2.63 |
| **AR(2)**        |    |    |    |    |    |
| US Data          | 2.53 | 2.88 | 3.03 | 3.14 | 3.13 |
| World Data       | 2.57 | 2.62 | 2.67 | 2.72 | 2.72 |
| **Smets-Wouters DSGE Bayesian** |    |    |    |    |    |
| US Data          | 2.79 | 2.95 | 2.89 | 2.80 | 2.71 |
| **Factor**       |    |    |    |    |    |
| US Data          | 2.24 | 2.48 | 2.50 | 2.67 | 2.86 |
| **RNN (Ours)**   |    |    |    |    |    |
| US Data          | 3.46 | 3.37 | 3.01 | 3.23 | 3.30 |
| World Data       | 2.35 | 2.52 | **2.50** | **2.62** | **2.60** |
| **AutoML (Ours)**|    |    |    |    |    |
| US Data          | 2.41 | 2.58 | 2.71 | 2.45 | 2.92 |
| **SPF Median**   |    |    |    |    |    |
| SPF Data         | 1.86 | 2.11 | 2.36 | 2.46 | 2.65 |

This table shows the results of our tested models. We compared the performance of all of the above models mentioned in this poster with our RNN and AutoML models. Our RNN with pooled data had the lowest RMSE (highest performance) for 3Q, 4Q, and 5Q ahead. The AutoML model trained with pooled data had the lowest RMSE for 1Q and 2Q ahead.