Explore FRED-QD: A Quarterly Database for Macroeconomic Research

Yifan Hao, Yueying Hu, Yakun Li, Yongliu Liu {yhaoah, yhucn, ylinv, yliuks}@connect.ust.hk

Department of Mathematics, HKUST

Introduction

Achieving robust forecasts for a single time series with many covariates and possible nonlinear effects is a problem worth investigating. In this paper, we first apply several manifold learning methods to do the dimension reduction of the FRED-QD dataset and select some relatively important macroeconomic indicators from hundreds of macroeconomic variables. Then we use factor-augmented regression (FR) to make an effective prediction and compare it with other forecasting methods. Detailed analysis is conducted.

Data Description

The FRED-QD dataset, developed by the Federal Reserve Bank of St. Louis, is a quarterly macroeconomic research database containing 234 economic time series. Here we consider the macroeconomic data from 1974 Q1 to 2021 Q4 as the overall sample and the sliding window length is set to H = 60. Thus, the testing set is the data from 1989 Q1 to 2021 Q4. Specifically, Figure 1 demonstrates the heat map of the aggregated quantile correlations between the 234 macroeconomic variables.

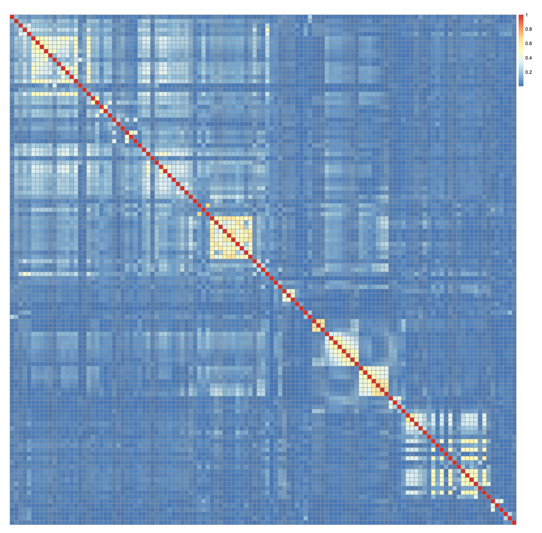

Fig 1: Hotplot for FRED-QD database.

Methodology

- **SPCA**[6] is short for Sparse Principal Component Analysis, which identifies and extracts sparse components from data, where only a small number of features contribute significantly to each component, allowing for a more interpretable and efficient representation of the data.
- **RPCA** [1] is short for Robust Principal Component Analysis, which separates data into low-rank and sparse components, effectively extracting the underlying structure while identifying and isolating outliers or noise.
- **MDS**[2] is short for Multidimensional Scaling, which projects the high-dimensional points to a low subspace by preserving the pairwise Euclidean distances of these points.
- **ISOMAP** [4] is short for Isometric Mapping, which is an extension of the classical MDS method by changing the Euclidean distance into the geodesic distance, which seeks a lower-dimensional embedding that maintains geodesic distances between all points.
- **LLE**[3] is short for Locally Linear Embedding, which attempts to project the data points into a low-dimensional space that best preserves the local geometric structures constructed in the neighborhoods of the data points
- **T-SNE** [5] is short for t-Distributed Stochastic Neighbor Embedding, which maps high-dimensional data into a lower one, preserving the local structure and emphasizing the separation of different clusters.

Results of embedding

Fig 2: Visualization of FRED-QD with dimensionality reduction techniques.

Results of forecasting

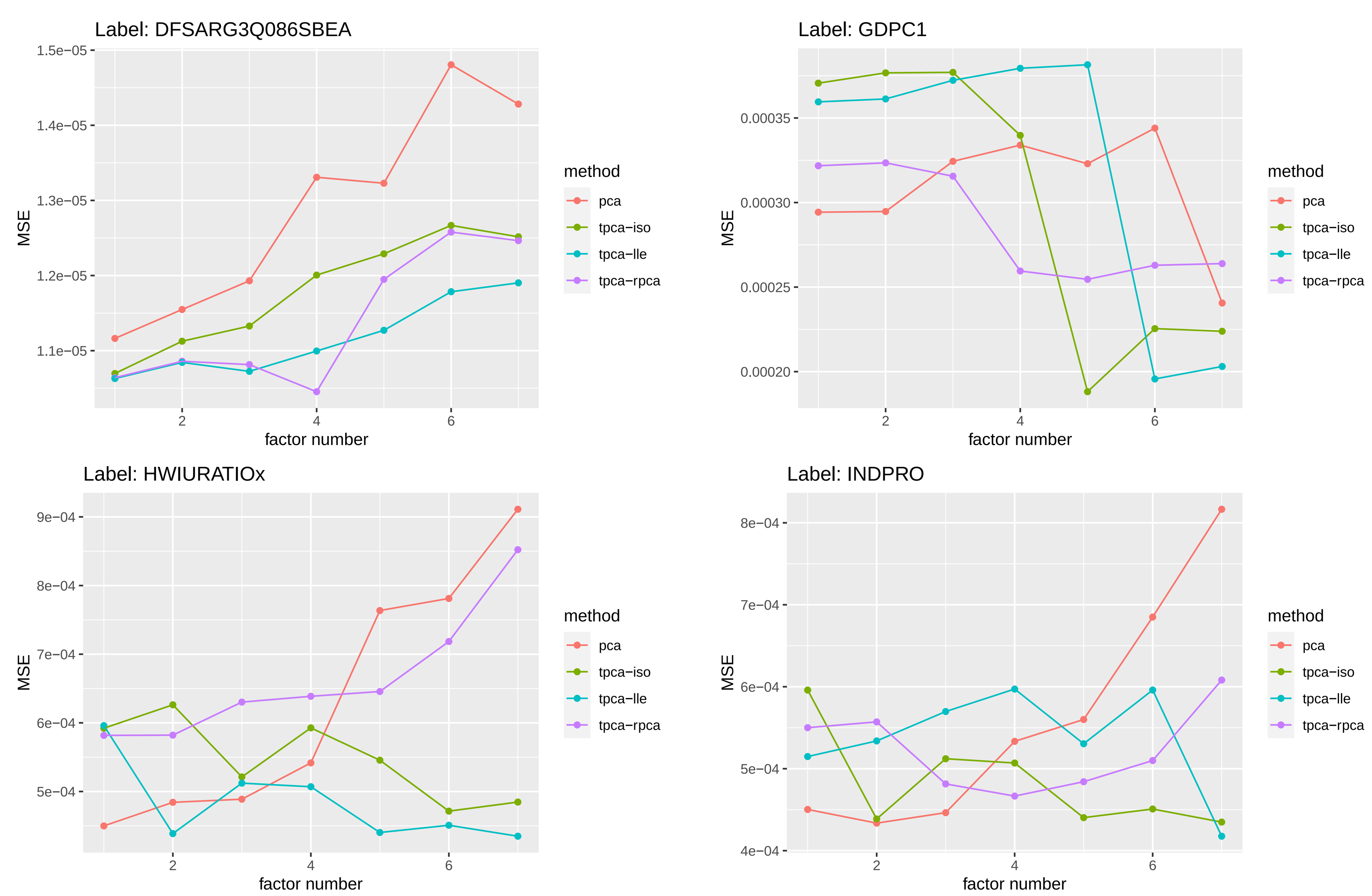

Fig 3: Visualization of forecasting via factor augmented model.

Taking forecasting on target variables with a fixed window width, we introduce the classical **Factor Augmented Regression** (FAR) model take forecasting. To be specific,

$$y_t = \alpha y_{t-1} + \boldsymbol{\beta}^T \boldsymbol{x}_t + \epsilon_t,$$

in which y_{t-1} is the lag term, and x_t is the multi-dimension factor, induced from original observers, as well as ϵ_t defined as noise.

However, capturing the factors from high dimensional space directly is not efficient, even not effective, due to the overfitting delima. Due to this, we prefer to reduce the data dimension first, then estimate the factors from such a low dimension space. And results above verify the high efficiency of this method.

Analysis and Conclusion

By applying the aforementioned dimension reduction techniques, we have selected several important economic indicators from a pool of 234 macroeconomic variables, including GDP, industrial production growth rate, employment, and unemployment rates. For each indicator, we have tested the out-of-sample forecasting performance of classical PCA method, as well as combing it with data reduction methods, using the remaining macroeconomic variables as covariates. The out-of-sample R2 results indicates that the FAR combing with general data reduction methods, generally exhibits superior predictive performance compared to the original PCA algorithm. As economic variables often exhibit heavy-tailed volatility and the relationship between their quantiles and covariates is not consistently constant, the FAR model combing with onolinear data reduction methods, which captures more infromation in different quantiles, consistently outperforms the other methods.

Furthermore, Fig. 3 displays the true and forecasted values of 'GDPC1' (Real GDP) and 'INDPRO' (Industrial Production Index) from 1989 Q1 to 2021 Q4 using different methods. The black dots represent the true observations, while the different colored lines depict the forecasts generated by each method. From the figure, we observe that all forecasting models capture the significant volatility in the time series, particularly during the subprime crisis in 2008 and the Covid-19 pandemic in 2020. These events profoundly impacted various economic indicators, including a sharp decline in GDP. Notably, the FR method exhibits more stationary forecasted series than the other methods and can recover to a steady state more quickly, showcasing its robustness.

Additionally, when predicting industrial production, only the FR method shows a notable increase in trend from 2020 Q2 to 2020 Q3, similar to the true trend. This suggests that the covariates already contain some information supporting industrial production growth at that particular moment, and the FR method is more effective in extracting and utilizing this information than the other methods. Therefore, it can be concluded that FR consistently delivers more robust performance in real-world scenarios.

References

[1] Emmanuel J Candès et al. "Robust principal component analysis?" In: *Journal of the ACM (JACM)* 58.3 (2011), pp. 1–37.

[2] AL Mead. "Review of the development of multidimensional scaling methods". In: *Journal of the Royal Statistical Society: Series D (The Statistician)* 41.1 (1992), pp. 27–39.

[3] Sam T Roweis and Lawrence K Saul. "Nonlinear dimensionality reduction by locally linear embedding". In: *science* 290.5500 (2000), pp. 2323–2326.

[4] Joshua B Tenenbaum, Vin de Silva, and John C Langford. "A global geometric framework for nonlinear dimensionality reduction". In: *science* 290.5500 (2000), pp. 2319–2323.

[5] Laurens Van der Maaten and Geoffrey Hinton. "Visualizing data using t-SNE." In: *Journal of machine learning research* 9.11 (2008).

[6] Hui Zou, Trevor Hastie, and Robert Tibshirani. "Sparse principal component analysis". In: *Journal of computational and graphical statistics* 15.2 (2006), pp. 265–286.

Contribution

- Report and analysis: Yueying Hu; Yakun Li.
- Code and model setting: Yifan Hao; Yonglin Liu.