University of Hawaii Economic Research Organization &
Department of Economics
University of Hawaiʻi at Manoa

Macroeconomic Forecasting in the Era of Big Data

Peter Fuleky

AUBER
Salt Lake City, Utah
October 14, 2018

Let X be a $mT \times K$ matrix containing observations on K variables over T years sampled at frequency m. Then

Tall data: $T \to \infty$ (long calendar span of data).

Wide data: $K \to \infty$ (large number of regressors).

Dense data: $m \to \infty$ (high-frequency intra-year sampling, regardless of whether the data are tall).

Exact (small) DFMs:

$$y_t = Mf_t + \epsilon_t$$
$$f_t = Tf_{t-1} + \eta_t$$
$$E(\epsilon_{it}\epsilon_{jt}) = 0 \text{ and } E(\eta_{it}\eta_{jt}) = 0, \text{ for } i \neq j$$

Approximate (large) DFMs:

$$y_t = \beta' F_t + \gamma(L) y_{t-1} + \epsilon_t$$
$$X_t = \Lambda F_t + \eta_t$$
$$\lim_{n \to 1} \sum_{i=1}^{n} \sum_{j=1}^{n} |E(\eta_{it}\eta_{jt})| < \infty$$

Uncorrelated F_t and η_t imply:

$$\Sigma_{XX} = \Lambda \Sigma_{FF} \Lambda' + \Sigma_{\eta\eta}$$

Estimate the factors and loadings via principal components (NLLS):

$$\min_{F_1 \dots F_T, \Lambda} \frac{1}{T} \sum_{t=1}^{T} (X_t - \Lambda F_t)'(X_t - \Lambda F_t) \quad \text{s.t.} \quad \Lambda'\Lambda = I$$

If $\Sigma_{\eta\eta}$ is non-diagonal and η is serially correlated use whitened X and weighted (generalized) principal components (GLS).

Choose the number of factors based on information criteria.

1. Estimate the factors, F_t, by principal components.
2. Estimate the factor loadings, Λ, from a regression of X_t on $\hat{F}_t$.
3. Estimate an autoregressive model for the residuals, $\hat{\eta}$, to obtain coefficients for idiosyncratic dynamics.
4. Estimate a vector autoregression of $\hat{F}_t$ on its lags, to obtain coefficients in the transition matrix.
5. Estimate the variance of the residuals of the vector autoregression.
6. Use the estimated coefficients in the Kalman smoother to obtain smoothed factors, $\tilde{F}_t$.
7. Project, y_t on $\tilde{F}_{t-h}$ to estimate β and make a direct forecast h steps ahead.

Mixed Frequency Data

UHERO
THE ECONOMIC RESEARCH ORGANIZATION AT THE UNIVERSITY OF HAWAIʻI

DFM The Kalman filter natively handles missing observations.

MIDAS Distributed lag models can be used to model relationships between low frequency and high frequency variables.

$$y_t^{(q)} = \beta_0 + \beta_1 b(L_m; \theta) x_{t+w-h}^{(m)} + \epsilon_t$$

or UMIDAS:

$$y_t^{(q)} = \delta_1 (L_m) x_{t+w-h}^{(m)} + \epsilon_t$$

Targeted Predictors

Before extracting factors from the data, narrow down X using:

Hard threshold retain X_i if its t statistics in individual regression of y on X_i exceeds threshold at significance level α.

Soft threshold based on penalized regressions:

(Ridge regression: $\min_{\beta} RSS + \lambda \sum_{j=1}^{N} \beta_j^2$)

Least absolute shrinkage selection operator: $\min_{\beta} RSS + \lambda \sum_{j=1}^{N} |\beta_j|$

Elastic net: $\min_{\beta} RSS + \lambda_1 \sum_{j=1}^{N} |\beta_j| + \lambda_2 \sum_{j=1}^{N} \beta_j^2$

Least angle regression: similar to forward stepwise regression.

Bagging Factors

Bootstrap Aggregation Algorithm

UHERO
THE ECONOMIC RESEARCH ORGANIZATION AT THE UNIVERSITY OF HAWAI'I

1. Extract the r largest common factors from the $T \times N$ matrix X of potential predictors.
2. Construct bootstrap samples $(y^*_{1+h}, \hat{f'}^*_1) \ldots (y^*_T, \hat{f'}^*_{T-h})$ by drawing with replacement blocks of m rows of the dataset.
3. Orthogonalize the bootstrap factor draws.
4. Estimate the loading coefficients, discard the insignificant factors, re-estimate the model, and predict $\hat{y}^*_{t+h}$.
5. The bagged forecast is the average of B bootstrap replications.

UHERO
THE ECONOMIC RESEARCH ORGANIZATION AT THE UNIVERSITY OF HAWAI'I

To forecast y_t consider the predictors $z_t = (Z_t, Z_{t-1} \dots Z_{t-pmax})$, with $Z_t = (y_{t-1}, F_{t1} \dots F_{tf}, F_{t1}^2 \dots F_{tr}^2)$.

Component-wise L_2 boost selects one predictor at a time.

1. Set $\hat{\Phi}_{t,0} = \overline{y}$ for each t.
2. For $m = 1 \dots M$:
   a. let $u_t = y_t - \hat{\Phi}_{t,m-1}$ be the current residuals;
   b. regress u_t on each z_i, select one that minimizes SSR;
   c. let $\hat{\phi}_m = z_{i_m^*} \hat{b}_{i_m^*}$;
   d. update $\hat{\Phi}_{t,m} = \hat{\Phi}_{t,m-1} + \nu \hat{\phi}_{t,m}$, where $0 < \nu \leq 1$.
3. The in-sample fit is $\hat{\Phi}_m(z) = \overline{y} + z' \hat{\beta}_m$, with recursion $\hat{\beta}_m = \hat{\beta}_{m-1} + \nu \hat{b}_m^*$, where $\hat{b}_m^* \neq 0$ only in the i^{th} position. At the final step, $\hat{\beta}_M$ will likely have many zero elements,

Block-wise L_2 boost selects the predictor and its lags jointly at each iteration.

Macroeconomic Forecasting
in the Era of Big Data - Table of Contents

UHERO
THE ECONOMIC RESEARCH ORGANIZATION AT THE UNIVERSITY OF HAWAI'I

Link: `http://www2.hawaii.edu/~fuleky/BigDataSite/index.html`

References I

Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221.

Bai, J. and Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2):304–317.

Bai, J. and Ng, S. (2009). Boosting diffusion indices. *Journal of Applied Econometrics*, 24(4):607–629.

Foroni, C. and Marcellino, M. G. (2013). A survey of econometric methods for mixed-frequency data. *Norges Bank Working Paper*.

Giannone, D., Reichlin, L., and Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 55(4):665–676.

Hirano, K. and Wright, J. H. (2017). Forecasting with model uncertainty: Representations and risk reduction. *Econometrica*, 85(2):617–643.

References II

Inoue, A. and Kilian, L. (2008). How useful is bagging in forecasting economic time series? a case study of us consumer price inflation. *Journal of the American Statistical Association*, 103(482):511–522.

Ng, S. (2014). Boosting recessions. *Canadian Journal of Economics/Revue canadienne d'économique*, 47(1):1–34.

Stock, J. and Watson, M. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20(2):147–162.

Stock, J. and Watson, M. (2006). Forecasting with many predictors. *Handbook of economic forecasting*, 1:515.

Stock, J. H. and Watson, M. W. (2011). Dynamic factor models. In *Oxford handbook on economic forecasting*. Oxford University Press.

Taieb, S. B. and Hyndman, R. J. (2014). Boosting multi-step autoregressive forecasts. In *ICML*, pages 109–117.