# Macroeconomic Data Transformation Matters

Philippe Goulet-Coulombe (UPenn), Maxime Leroux (UQÀM), Dalibor Stevanovic (UQÀM) and **Stéphane Surprenant** (UQÀM)

October 2020

# Content

- Accuracy gains from machine learning are well documented. **Some recent examples** : Kim and Swanson (2018), Goulet-Coulombe *et al.* (2019) and Meideiros *et al.* (2019), among others ;
- Typical data transformations in time series used to remove low frequency movements may be suboptimal for machine learning methods ;
- Deep neural networks may automate data transformations, but macroeconomic samples are short and noisy making manual feature engineering advisable (Kuhn and Johnson, 2019) ;
- Moreover, careful feature engineering can encode **prior knowledge** and help improve forecasting accuracy.

**Goals** :

- Propose rotations of original data which helps encode "time series friendly" priors into machine learning models ;
- And, of course, we seek to compare the performances of several data transformations and the combinations thereof ;
- Note that we focus on the **transformations of the predictors**. Throughout, targets are defined as

$$y_t^{(h)} = \frac{1}{h}(ln(Y_t) - ln(Y_{t-h}))$$

where $h$ is the forecasting horizon.

- **Model** : $y_{t+h} = g\left(f_Z(H_t)\right) + \epsilon_{t+h}$
- **Objective** : $\min_{g \in \mathcal{G}} \left\{ \sum_{t=1}^{T} \left( y_{t+h} - g\left(f_Z(H_t)\right)^2 + \text{pen}(g, \tau) \right) \right\}$
- $H_t$ is the data, $f_Z$ is the feature engineering step, $g$ is the model and pen is a penalty function with hyperparameter vector $\tau$. Hence $Z_t := f_Z(H_t)$ would be the feature matrix.
- **Forecast error decomposition** :

$$y_{t+h} - \hat{y}_{t+h} = \underbrace{g^*(f_Z^*(H_t)) - g(f_Z(H_t))}_{\text{approximation error}} + \underbrace{g(Z_t) - \hat{g}(Z_t)}_{\text{estimation error}} + e_{t+h}.$$

- We focus on how our choice of $f_Z$ (transformations or combinations thereof) impacts forecast accuracy.

We consider "older" or more common candidates :

- **X** : Differentiate the data in levels or logarithms.
- **F** : PCA estimates of linear latent factors of **X** as in Stock and Watson (2002a,b) and Bai and Ng (2008).
- **H** : (Log-)Level of the series.

We consider "newer" or less common candidates :

- **MARX** (Moving average rotation of $\mathbf{X}$) : We use order $p = 1, ..., P_M$ moving averages of each variable in $X$. This is motivated by Shiller (1973). The following model :

$$y_t = \sum_{p=1}^{p=P} X_{t-p}\beta_p + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2)$$

$$\beta_p = \beta_{p-1} + u_p, \quad u_p \sim N(0, \sigma_u^2 I_K).$$

  This can be estimated by a Ridge regression which may be parametrized using inputs $Z := XC$ as inputs and where $C = I_K \otimes c$ with $c$, a lower triangular matrix of ones.

- This transformation implicitly shrinks $\beta_p$ to $\beta_{p-1}$.

- **MAF** (Moving average factors) : Let $\tilde{X}_{t,k}$ be the $k$-th variable lag matrix defined as

$$\tilde{X}_{t,k} = \left[X_{t,k}, LX_{t,k}, \ldots, L^{P_{MAF}} X_{t,k}\right]$$
$$\tilde{X}_{t,k} = M_t \Gamma'_k + \tilde{\epsilon}_{k,t}.$$

  We estimate $M_t$ by PCA and use the same number of factors for all variables.

- This is related to Singular Spectrum Analysis – except that SSA would use the whole common component instead of focusing on latent factors.

## Tableau – Summary : Feature Matrices and Models

| Transformations | Feature Matrix |
|---|---|
| F | $Z_t^{(F)} := [F_t, LF_t, \ldots, L^{p_f} F_t]$ |
| X | $Z_t^{(X)} := [X_t, LX_t, \ldots, L^{p_X} X_t]$ |
| MARX | $Z_t^{(MARX)} := \left[MARX_{1t}^{(1)}, \ldots, MARX_{1t}^{(p_{MARX})}, \ldots, MARX_{Kt}^{(1)}, \ldots, MARX_{Kt}^{(p_{MARX})}\right]$ |
| MAF | $Z_t^{(MAF)} := \left[MAF_{1t}^{(1)}, \ldots, MAF_{1t}^{(r_1)}, \ldots, MAF_{Kt}^{(1)}, \ldots, MAF_{Kt}^{(r_K)}\right]$ |
| Level | $Z_t^{(Level)} := [H_t, LH_t, \ldots, L^{p_H} H_t]$ |
| **Model** | **Functional space** |
| Autoregression (AR) | Linear |
| Factor Model (FM) | Linear |
| Adaptive Lasso (AL) | Linear |
| Elastic Net (EN) | Linear |
| Linear Boosting (LB) | Linear |
| Random Forest (RF) | Nonlinear |
| Boosted Trees (BT) | Nonlinear |

- Several combinations of many of those transformations are also considered.

## The Horse Race

- **Direct forecasts**
- **Data** : FRED-MD
- **Targets** : Industrial production, non farm employment, unemployment rate, real personal income, real personal consumption, retail and food services sales, housing starts, M2 money stock, consumer and producer price index
- **Horizons** : $h \in [1, 3, 6, 9, 12, 24]$
- **POOS Period** : 1980M1-2017M12
- **Estimation Window** : Expanding from 1960M1

## Tableau – Best Specifications in Terms of MSE

| | INDPRO | EMP | UNRATE | INCOME | CONS | RETAIL | HOUST |
|---|---|---|---|---|---|---|---|
| H=1 | RF (MARX, X, L) | RF (F, MARX, X, L) | LB (F, MARX, X, L) | BT (F, MARX, X) | FM (F) | FM (F) | FM (F) |
| H=3 | RF (MARX) | RF (F, MARX, X) | RF (F, MARX) | RF (MARX) | FM (F) | AL (X, MAF) | BT (F, X, MAF) |
| H=6 | AL (F) | RF (F, MARX, X, L) | LB (F) | RF (F, MARX) | RF (F, X, L) | AL (F, MARX) | BT (X, L) |
| H=9 | LB (F) | LB (F, X, L) | EN (F) | RF (MARX, X) | RF (F, X, L) | AL (MARX, X) | RF (F) |
| H=12 | RF (MAF) | LB (X, L) | RF (F, MAF) | RF (F) | RF (F, X, L) | RF (F, X, MAF) | RF (F) |
| H=24 | RF (X, L) | LB (MARX, X, L) | RF (F) | RF (F) | RF (F) | EN (X, L) | RF (F, X) |

| | M2 | CPI | PPI |
|---|---|---|---|
| H=1 | BT (F, X, L) | EN (MARX) | EN (MARX) |
| H=3 | BT (F, X, L) | RF (F) | EN (MARX, X, L) |
| H=6 | BT (X, L) | RF (F) | RF (F) |
| H=9 | BT (X, L) | RF (F) | RF (F) |
| H=12 | BT (X, L) | EN (F) | RF (F) |
| H=24 | RF (F) | AL (F) | RF (F) |

Note : Bullet colors represent data transformations included in the best model specifications : *F*, *MARX*, *X*, *L*, *MAF*.

- The **marginal contribution** of a transformation to model performance can be evaluated using a panel regression :

$$R^2_{t,h,v,m} = \alpha_f + \psi_{t,v,h} + v_{t,h,v,m}$$

where

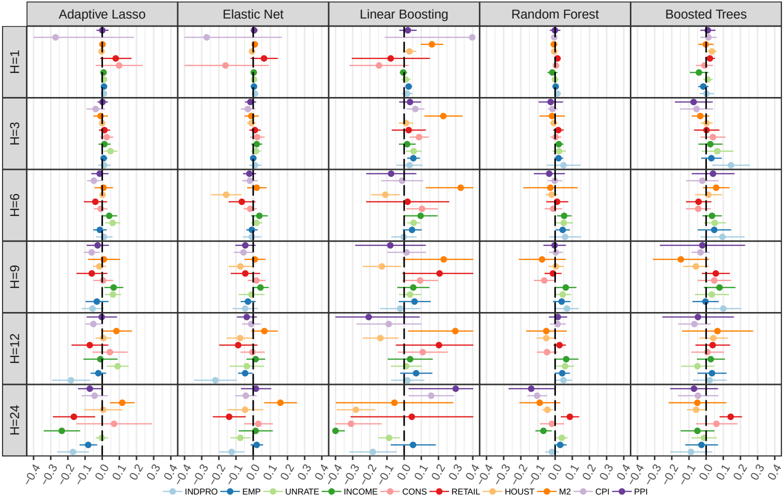
- $R^2_{t,h,v,m} := 1 - \frac{\epsilon^2_{t,h,v,m}}{\frac{1}{T}\sum_{t=1}^{T}\left(y^{(h)}_{v,t} - \bar{y}^{(h)}_{v}\right)^2}$
- $\psi_{t,v,h}$ are time, variable and horizon fixed effects
- $\epsilon_{t,h,v,m}$ is the time $t$, horizon $h$, variable $v$ and model $m$ forecast error
- $\alpha_f$ is one of $\alpha_{MARX}, \alpha_{MAF}, \alpha_F$ associated with the corresponding transformations. The null hypothesis is $\alpha_f = 0$.

## Main Findings

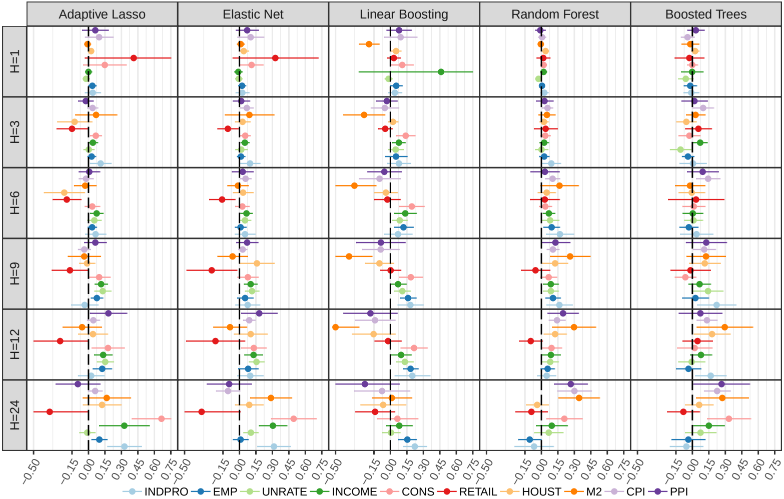
- MARX is especially potent when used in combination with nonlinear models to forecast measures of real activity. It's especially true around recessions Cumulative Squred Errors (Real).
- Factors helps a lot with Random Forest and Boosted Trees, especially at longer horizons, in line with results in Goulet-Coulombe *et al.* (2019).
- Random Forests with factors gained a lot of ground for year-ahead forecasting Cumulative Squred Errors (CPI).
- MAF regressions focus on models which includes **X**. Results are more muted, but it seems to help Random Forests and Linear Boosting for horizons of 6 and 9 months.

## Conclusion

- At shorter horizons, combining non-standard and standard data transformations helps reduce RMSE.
- MARX is especially potent when used in combination with nonlinear models to forecast measures of real activity, especially around recessions.
- Factors remain one of the most effective feature engineering tool available for macroeconomic forecasting, even for inflation.

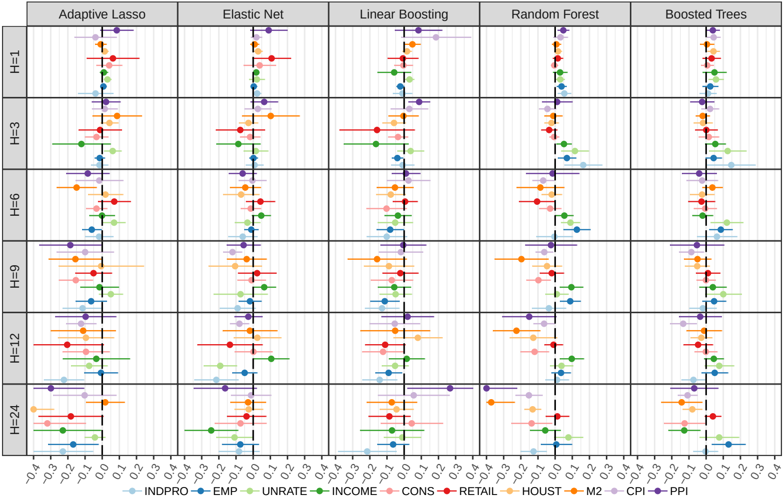



Back FIGURE – **MAF** Average Treatment Effects ((h,v) Subsets)




Back

FIGURE – F Average Treatment Effects ((h,v) Subsets)




Back FIGURE – **MARX** Average Treatment Effects ((h,v) Subsets)

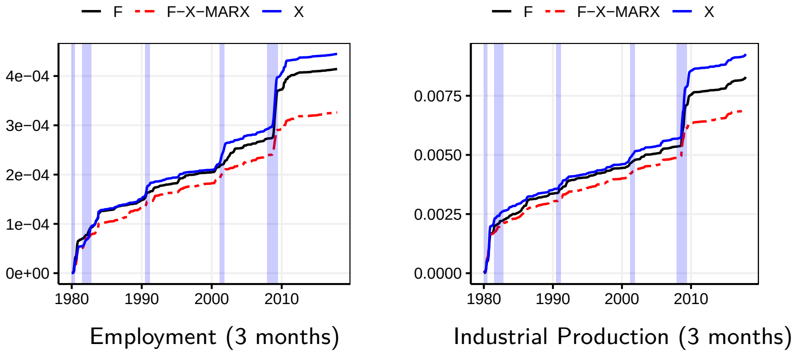



FIGURE – Cumulative Squared Errors (Random Forest)

Back

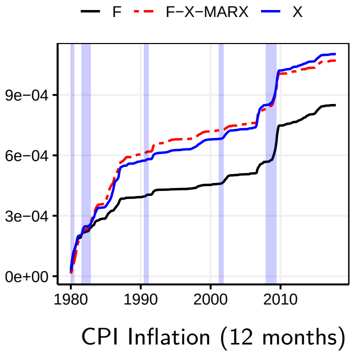
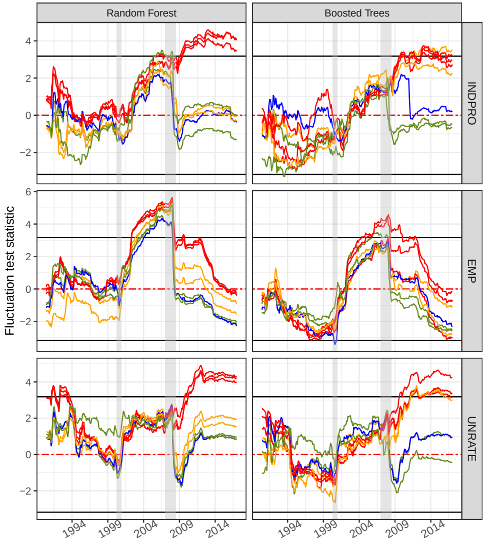
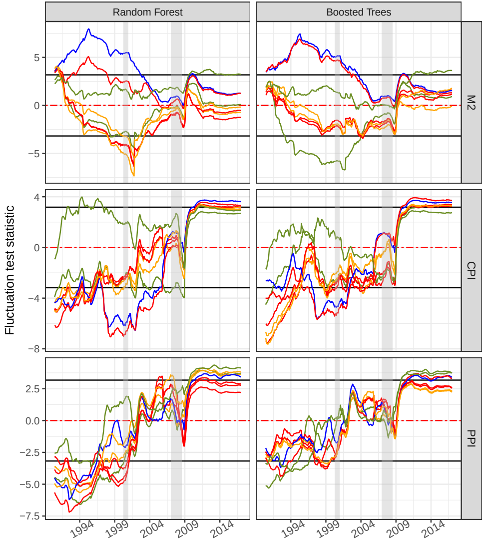



FIGURE – Cumulative Squared Errors (Random Forest)

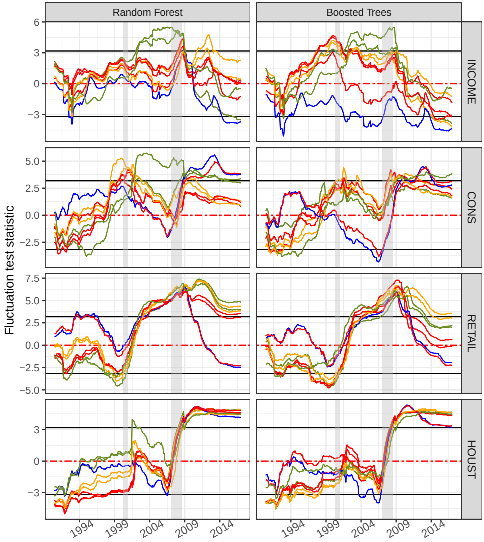
Back




**Notes :** Giaconomini-Rossi fluctuation tests for 3 months at 10%. Improvements lie above the upper line. Transformations : *F,F-X*, *F-MARX,F-X-MARX,F-X-MARX-Level*, *F-X-Level*, *F-MAF,F-X-MAF*.




**Notes :** Giaconomini-Rossi fluctuation tests for 12 months at 10%. Improvements lie above the upper line. Transformations : *F,F-X*, *F-MARX,F-X-MARX,F-X-MARX-Level*, *F-X-Level*, *F-MAF,F-X-MAF*.




**Notes :** Giaconomini-Rossi fluctuation tests for 12 months at 10%. Improvements lie above the upper line. Transformations : *F,F-X*, *F-MARX,F-X-MARX,F-X-MARX-Level*, *F-X-Level*, *F-MAF,F-X-MAF*.