Sparse Warcasting

Forecasting in a data-rich but statistics-poor environment

Mihnea Constantinescu

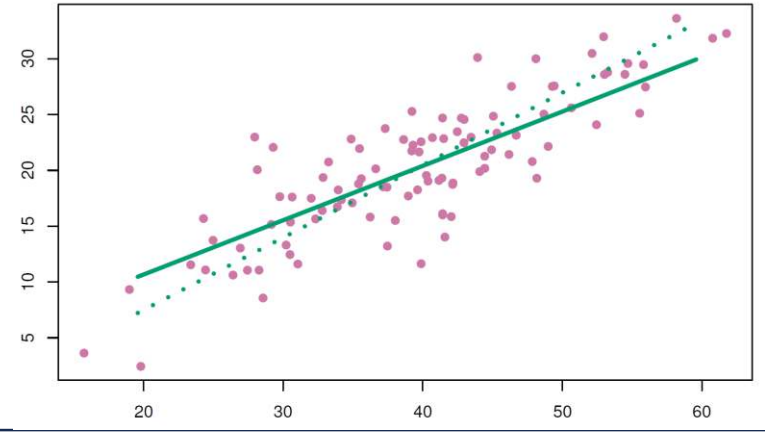
National Bank of Ukraine & UvA

11th BCC Conference

September 2023

Traditional Dynamic Factor Models

- Dynamic Factor Models are estimated using **Principal Components,** an unsupervised factor extraction method
- **PCR** factors, a dimensionality reduction technique, is an unsupervised method (PC factors - dotted)
- **Forecasting/nowcasting** is a supervised application (PLS factors- solid)



$$
\begin{aligned}
x_t &= \Lambda f_t + \xi_t, & \xi_t &\sim \mathbb{N}(0, \Sigma_\xi) \\
f_t &= \sum_{i=1}^{p} A_i f_{t-i} + \zeta_t, & & \\
\zeta_t &= B\eta_t, & \eta_t &\sim \mathbb{N}(0, I_q) \\
\widehat{y}_t &= \beta' f_t. & &
\end{aligned}
$$

Supervised DFM – The Algorithms

- PCR, extract factors to be independent of each other **regardless** of who **y** is.

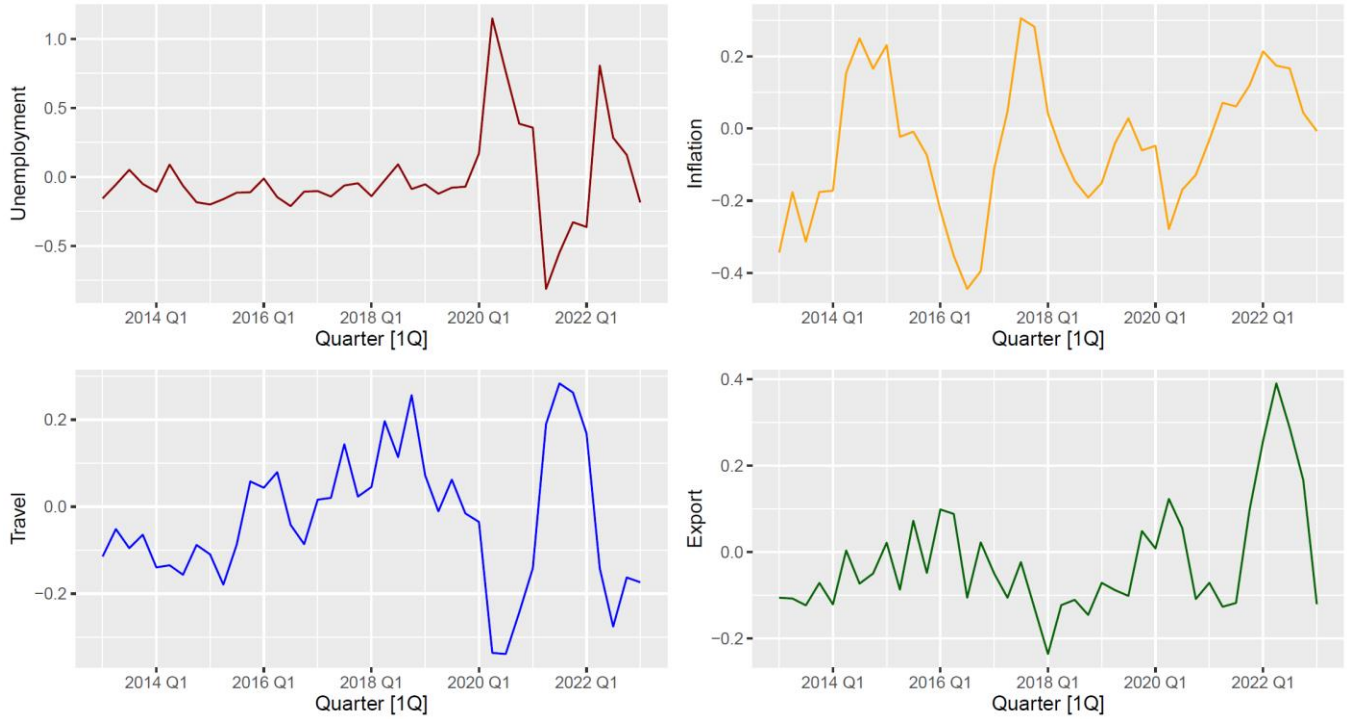
$$\max_\alpha \mathrm{Var}(\mathbf{X}\alpha)$$
$$\text{subject to } ||\alpha|| = 1,\ \alpha^T \mathbf{S} v_\ell = 0,\ \ell = 1, \ldots, m-1,$$

- PLS, being supervised by **y**, identifies the components or factors to be independent of each other while also having high correlation with the target **y** {Wold et al. 1984}

$$\max_\alpha \mathrm{Corr}^2(\mathbf{y}, \mathbf{X}\alpha)\mathrm{Var}(\mathbf{X}\alpha)$$
$$\text{subject to } ||\alpha|| = 1,\ \alpha^T \mathbf{S} \hat{\varphi}_\ell = 0,\ \ell = 1, \ldots, m-1.$$

Data

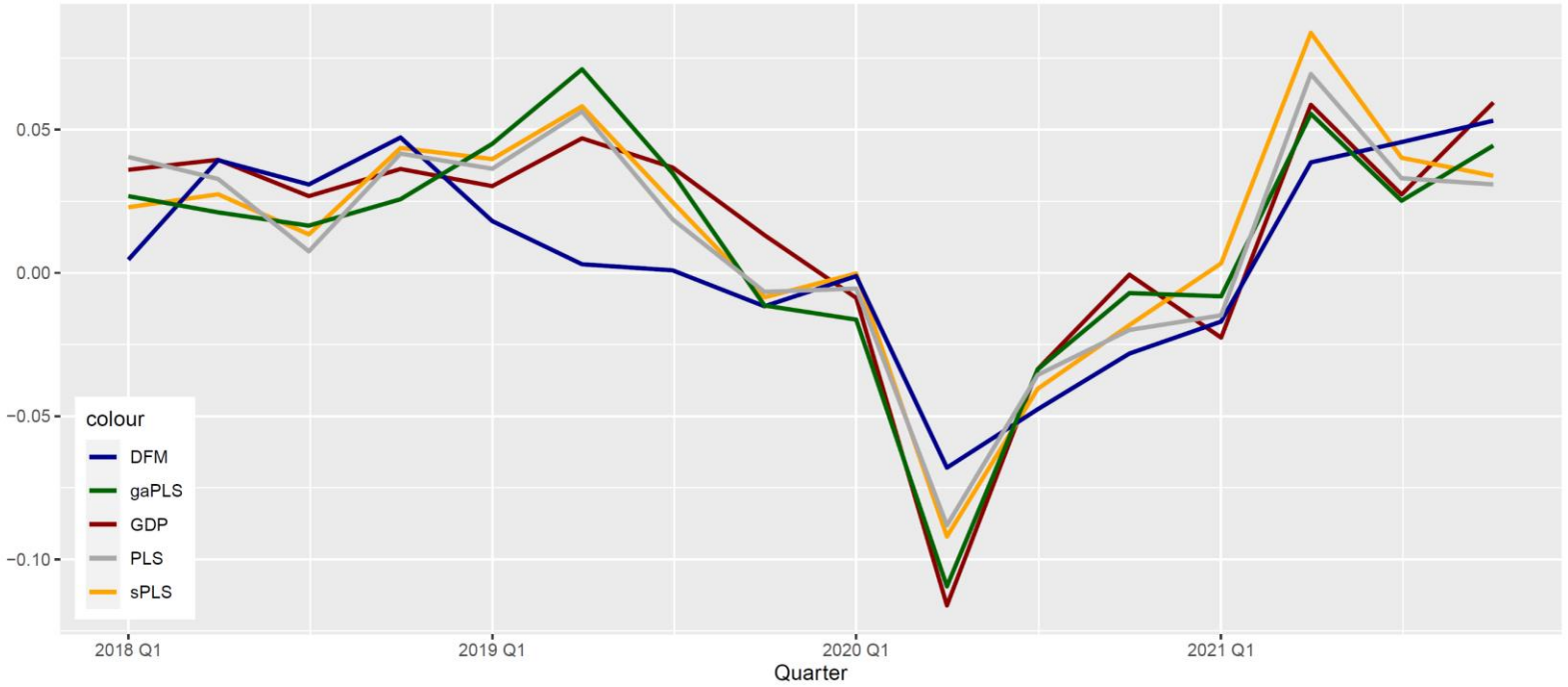
- **Input**: A list of Google Trends time-series: Topics and Keywords
- **Target**: quarterly deflated y-o-y GDP from 2012 to 2021
- **Objective**: Nowcast and near-forecast GDP in the early stages of the Russian 2022 invasion



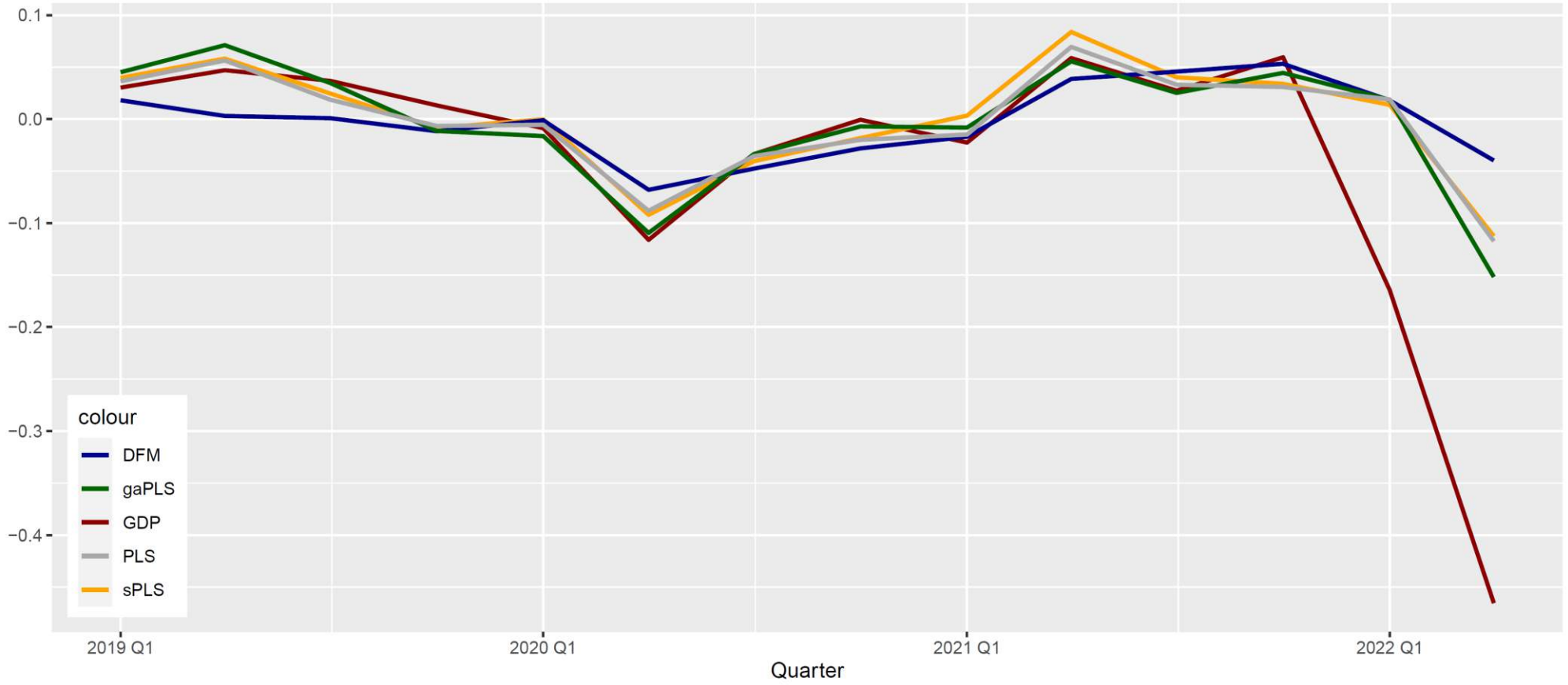
Methodology & In-sample Fit

- Variable selection is performed by a Genetic Algorithm, several alternative also considered
  - Sparse PLS by Chun & Keles and Filter methods

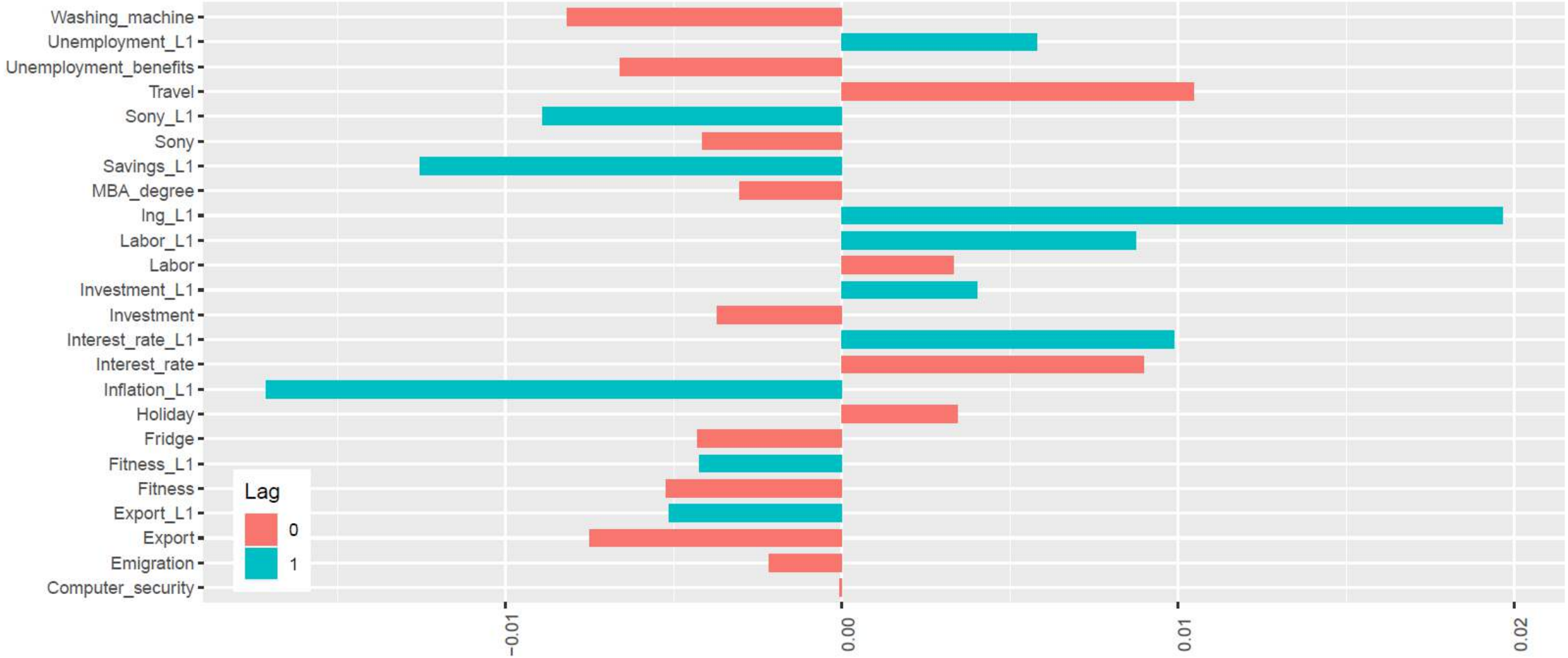
- The algorithm considered **the joint problem** of finding relevant variables & the appropriate number of latent factors



Out-of-sample



Selected Variables



Conclusions

- Preliminary results show the value of using supervised algorithms in conjunction with variable selection methods to nowcast and near-term forecast when no official statistics are available

- PLS outperforms PCR primarily thanks to its better identification of latent factors
  - Lower number of factors also relevant given the small sample