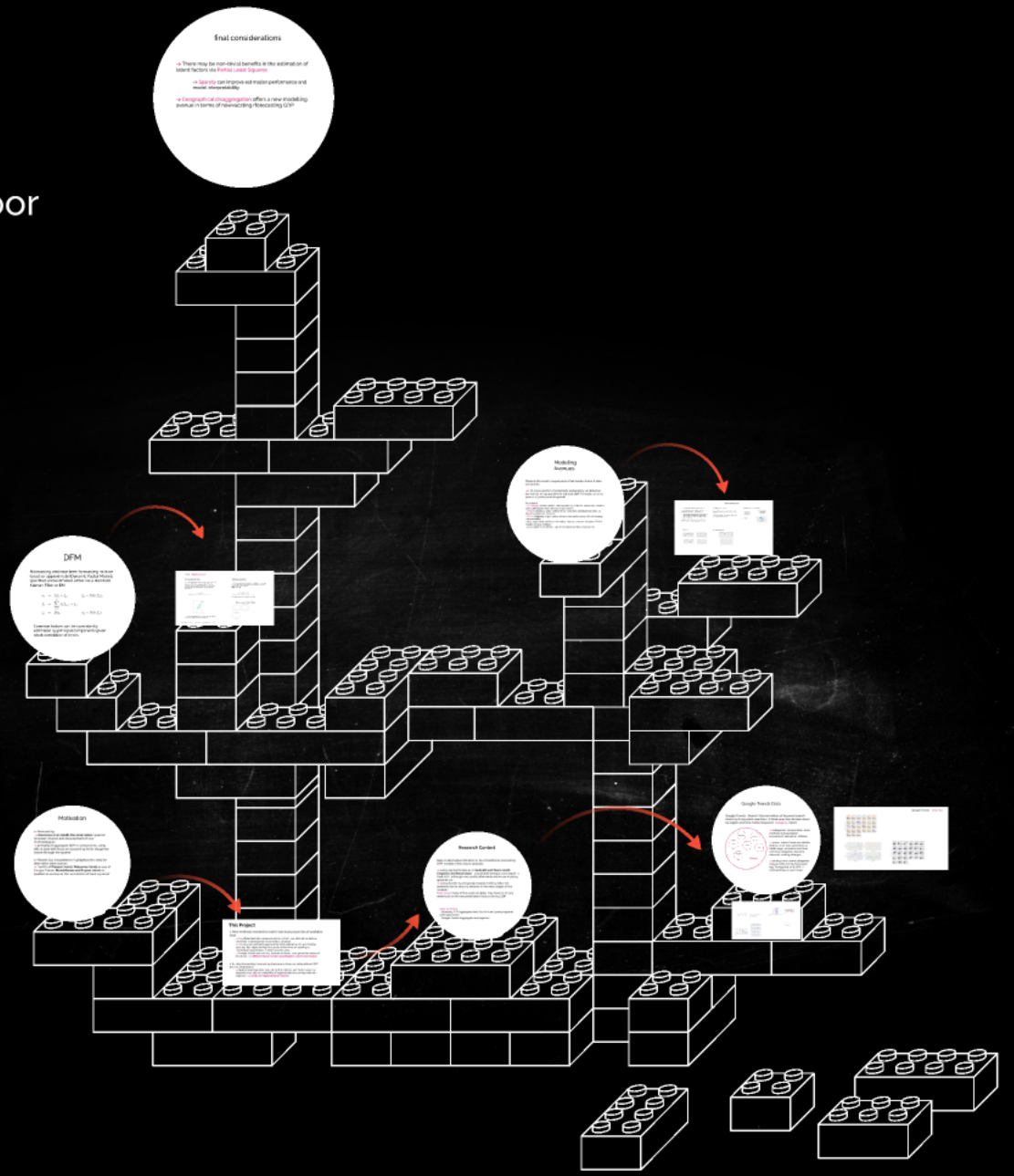
Sparse Warcasting

Forecasting in a data-rich but statistics-poor environment



Mihnea Constantinescu
National Bank of Ukraine and
University of Amsterdam

June 2023

Motivation

>> Nowcasting
-> **Giannone et al. (2008); Doz et al. (2011)** spurred renewed interest and development of new methodologies
-> primarily of aggregate GDP or components, using official data with focus on accounting for its staggered nature through the quarter

>> Recent Covid experience highlighted the need for alternative data sources
-> VoxEU of **Diebold (2020); Woloszko (2020)** on use of Google Trends; **Blanchflower and Bryson (2021)** on qualitative surveys or the "economics of walking about"

Motivation

>> Nowcasting
-> **Giannone et al. (2008); Doz et al. (2011)** spurred renewed interest and development of new methodologies
-> primarily of aggregate GDP or components, using official data with focus on accounting for its staggered nature through the quarter

>> Recent Covid experience highlighted the need for alternative data sources
-> VoxEU of **Diebold (2020); Woloszko (2020)** on use of Google Trends; **Blanchflower and Bryson (2021)** on qualitative surveys or the "economics of walking about"

>> The Feb 24th russian invasion led to a freeze of all official data gathering by local and national statistical agencies
--> only left with alternative data sources

This Project

1. New methods needed to match statistcal properties of available data

   -> No official statistics released until mid 2022; use alternative data as identified in development economics literature

   -> Lit considers primarily peace-time GDP estimation: NL and Twitter possibly "flip" signs during the course of the invasion leading to inconsistent parameters; Twitter data for a fee

   -> Google Trends not entirely immune to these issue given the nature of the shock **->> different factor model specification and/or estimation**

2. Existing forecasting/nowcasting frameworks focus on national level GDP and/or components

   -> Spatial heterogeneity: replicating the national estimation exercise depends crucially on availability of regional data (assuming national = regional) **->> scope for regional factor model**

Research Context

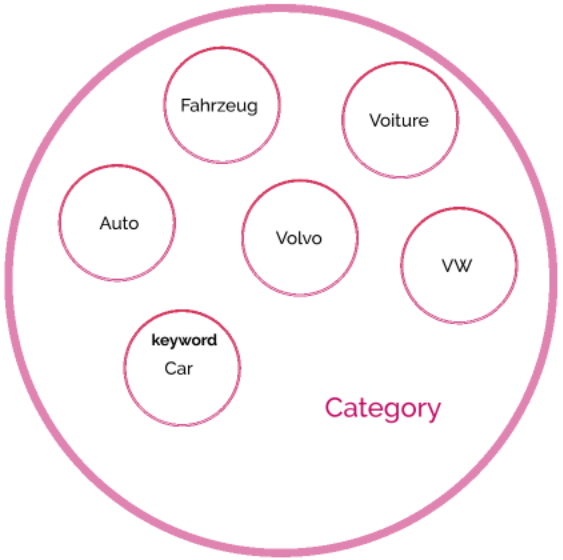
Rely on alternative indicators to build traditional nowcasting DFM models of key macro variables

-> using payments data as in **Galbraith and Tkacz (2018)**, **Chapman and Desai (2021)** : unavailable timespan and depth in April 2022 although very useful afterwards and in use in policy applications
-> using electricity and google/apple mobility data: not available due to security reasons in the early stages of the invasion;
Main issue: many of the used variables may have no or very weak load on the assumed latent factors driving GDP

data-summary:
- Quarterly GDP aggregate data (Q4 2021) and yearly regional GDP data (2020)
- Google Trends (aggregate and regional)

Google Trends Data

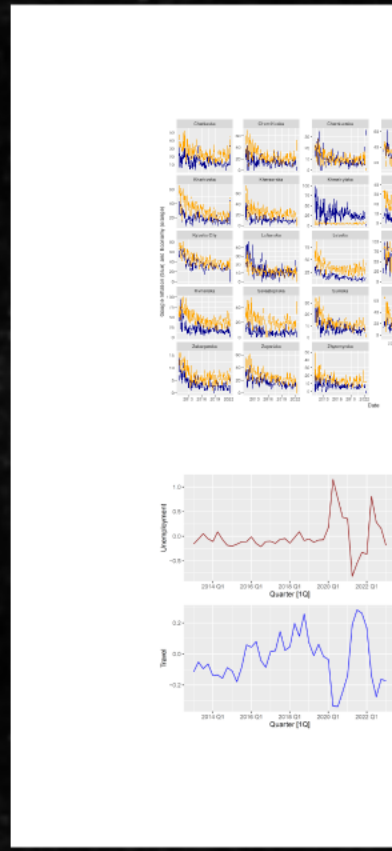
Google Trends = Search Volume indices of keyword search intensity (# keyword searches/ # total searches) broken down by region and time frame {Keyword, Category, Topic}



-> categories: consumption, labor markets, transportation, investment, education, inflation

-> some issues: these are relative indices (w.r.t. total searches); as total usage increases over time and new categories become relevant, ranking changes.

-> during a war, some categories may give the wrong impression (eg. Transportation & GDP in normal times vs. war times)



First Econ Applications

> Ettredge et al. (2005) is one of the earlier references using Google Search activity to forecast the US unemployment rate.
> Askits Zimmermann (2009) "Google Econometrics and Unemployment Forecasting" use google searches related to unemployment to forecast offical figures several months ahead.
> This is particularly relevant as in 2008-2009, data releases on key macrovariables are usually delayed several months as compared to observed macro and financial shocks

> Choi and Varian 2010/2011 "Prediciting the Present with Google Trends" makes a strong point in the potential use of google trends data to nowcast a multitude of economic variables such as automobile sales, unemployment claims,travel destination planning, and consumer confidence.

>Wu and Brynjolfsson (2010) leverage Google Search Data to forecast house prices

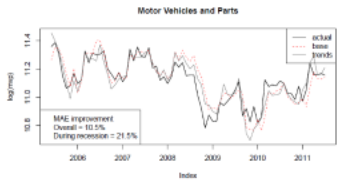
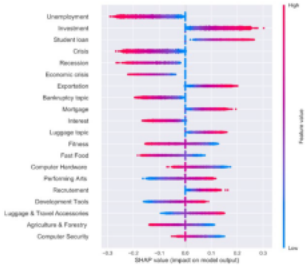


Figure 2: Motor Vehicles and Parts

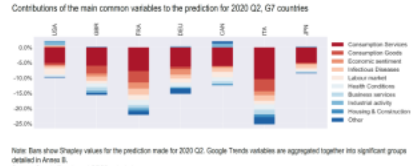
Most are time-series models: AR(n) and deep-learnig

Let y_t be the log of the observation at time t. We first estimate a simple baseline seasonal AR-1 model $y_t = b_1 y_{t-1} + b_{12} y_{t-12} + e_t$ for the period 2004-01-01 to 2011-07-01.

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.67266    0.76355   0.881 0.381117
lag(y, -1)   0.64345    0.07332   8.776 3.59e-13 ***
lag(y, -12)  0.29565    0.07282   4.060 0.000118 ***
---
Multiple R-squared: 0.7185,Adjusted R-squared: 0.7111
```



Note: Shapley values are the contributions of a variable to the GDP growth estimate predicted by the model. Variables are ranked by importance, and for each variable. Each point correspond to an observation (that is a given month * a given country) and its colour depends on the value of the variable.
Source: OECD calculations



Note: Bars show Shapley values for the prediction made for 2020 Q2. Google Trends variables are aggregated together into significant groups detailed in Annex B.
Source: Google Trends and OECD calculations.

> During the COVID period, timeliness and depth of the Google Trends data became essential to nowcast the speed and severity of the economic contraction.

Woloszko (2020) and Burri Kaufmann (2021) build nowcasting models with search data volumes as key high frequency inputs (all other hard & soft series available)

Indicator	Type	Frequency	Release	Relationship to GDP
GDP	Hard	Quarterly (monthly for GBR, CAN and SWE)	Usually 1-2 months after the end of the quarter	
Industrial production	Hard	Monthly	Around 30-55 days after the end of the month	Linear
Retail sales	Hard	Monthly	Around 8-10 weeks after the end of the month	Linear
PMIs	Soft	Monthly	Around start of the next month	Linear in normal times, non-linear around crises
Consumer confidence	Soft	Monthly	Around start of the next month	Linear in normal times, non-linear around crises
Google Mobility	High-frequency	Daily	With a 7-day delay	Difficult to calibrate as historical data start mid-February 2020.
Google Trends	High-frequency	Daily, Weekly or Monthly	With a 5-day delay	Model-based relationship

Source: OECD.

> Burri Kaufmann (2021) use daily financial data to build a high-frequency GDP tracker for Switzerland; the methods build on the availability of deep financial markets, a chimera for many developing economies

> A two-factor model is estimated, tracking CH and non-CH variables

https://github.com/dankaufmann/f-curve

$$[X\ X^*] = [f\ f^*]\begin{bmatrix}\lambda_{11} & 0\\ \lambda_{21} & \lambda_{22}\end{bmatrix} + e$$

where X, X^* denote the data matrices comprising domestic and foreign variables, respectively. In addition f, f^* represent the domestic and foreign factors and $\lambda_{11}, \lambda_{21}, \lambda_{22}$ are the loading matrices.

$$y_{\tau+h} = \alpha_h + \beta_{h,1} f_{\tau|t} + \beta_{h,2} f_{\tau-1} + \nu_{\tau+h}$$

> Ettredge et al. (2005) is one of the earlier references using Google Search activity to forecast the US unemployment rate.
> Askits Zimmermann (2009) "Google Econometrics and Unemployment Forecasting" use google searches related to unemployment to forecast offical figures several months ahead.
> This is particularly relevant as in 2008-2009, data releases on key macrovariables are usually delayed several months as compared to observed macro and financial shocks

> Choi and Varian 2010/2011 "Prediciting the Present with Google Trends" makes a strong point in the potential use of google trends data to nowcast a multitude of economic variables such as automobile sales, unemployment claims,travel destination planning, and consumer confidence.

>Wu and Brynjolfsson (2010) leverage Google Search Data to forecast house prices

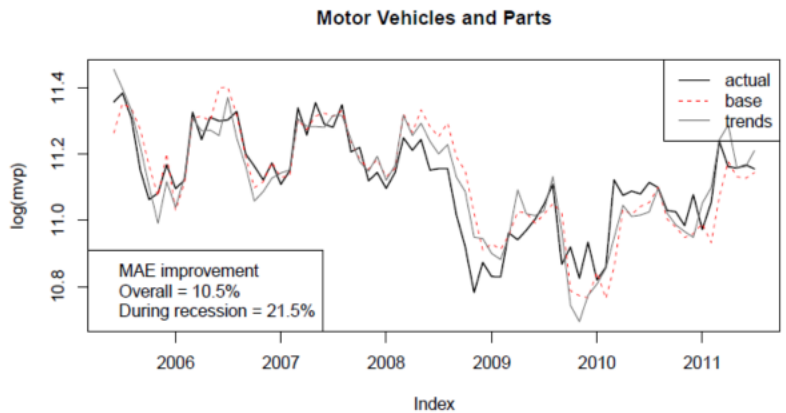


Figure 2: Motor Vehicles and Parts

Most are time-series models; AR(n) and deep-learnig

Let y_t be the log of the observation at time t. We first estimate a simple baseline seasonal AR-1 model $y_t = b_1 y_{t-1} + b_{12} y_{t-12} + e_t$ for the period 2004-01-01 to 2011-07-01.

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.67266    0.76355   0.881 0.381117
lag(y, -1)   0.64345    0.07332   8.776 3.59e-13 ***
lag(y, -12)  0.29565    0.07282   4.060 0.000118 ***
---
Multiple R-squared: 0.7185,Adjusted R-squared: 0.7111
```

> During the COVID period, timeliness and depth of the Google Trends data became essential to nowcast the speed and severity of the economic contraction.

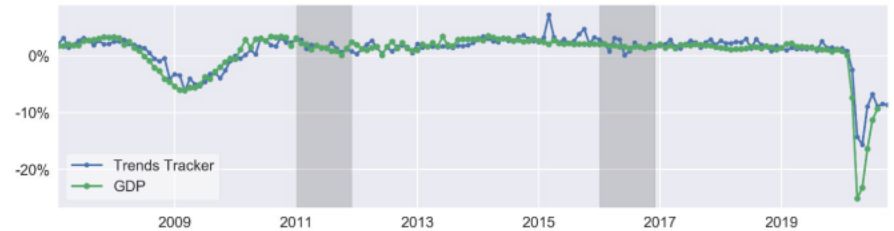
Woloszko (2020) and Burri Kaufmann (2021) build nowcasting models with search data volumes as key high frequency inputs {all other hard & soft series available}

Indicator	Type	Frequency	Release	Relationship to GDP
GDP	Hard	Quarterly (monthly for GBR, CAN and SWE)	Usually 1-2 months after the end of the quarter	
Industrial production	Hard	Monthly	Around 30-55 days after the end of the month	Linear
Retail sales	Hard	Monthly	Around 8-10 weeks after the end of the month	Linear
PMIs	Soft	Monthly	Around start of the next month	Linear in normal times, non-linear around crises
Consumer confidence	Soft	Monthly	Around start of the next month	Linear in normal times, non-linear around crises
Google Mobility	High-frequency	Daily	With a 7-day delay	Difficult to calibrate as historical data start mid-February 2020.
Google Trends	High-frequency	Daily, Weekly or Monthly	With a 5-day delay	Model-based relationship

Source: OECD.

Figure 2. Nowcasting GDP growth with Google trends (M-1 forecast) (*contd.*)

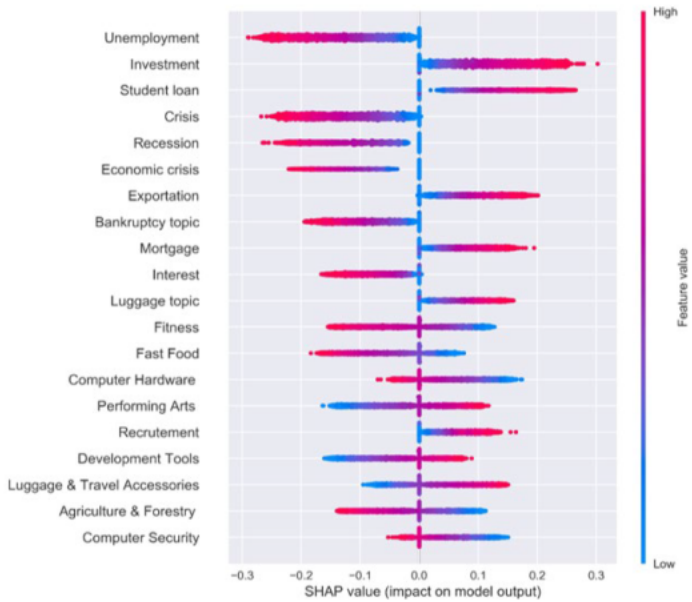
Panel B. United Kingdom



Panel C. Spain

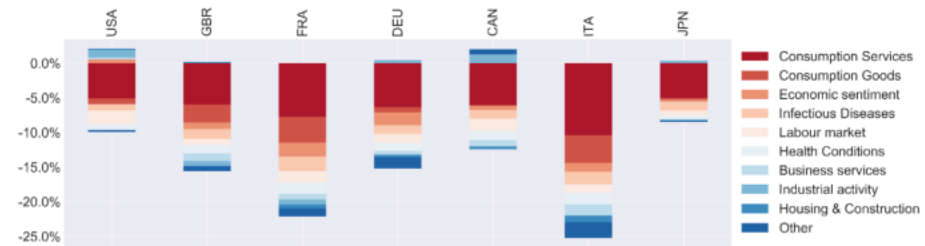
Panel D. Italy

Panel E. Germany



Note: Shapley values are the contributions of a variable to the GDP growth estimate predicted by the model. Variables are ranked by importance, and for each variable. Each point correspond to an observation (that is a given month * a given country) and its colour depends on the value of the variable.
Source: OECD calculations

Contributions of the main common variables to the prediction for 2020 Q2, G7 countries



Note: Bars show Shapley values for the prediction made for 2020 Q2. Google Trends variables are aggregated together into significant groups detailed in Annex B.
Source: Google Trends and OECD calculations.

> Burri Kaufmann (2021) use daily financial data to build a high-frequency GDP tracker for Switzerland; the methods build on the availability of deep financial markets, a chimera for many developing economies

> A two-factor model is estimated, tracking CH and non=CH variables

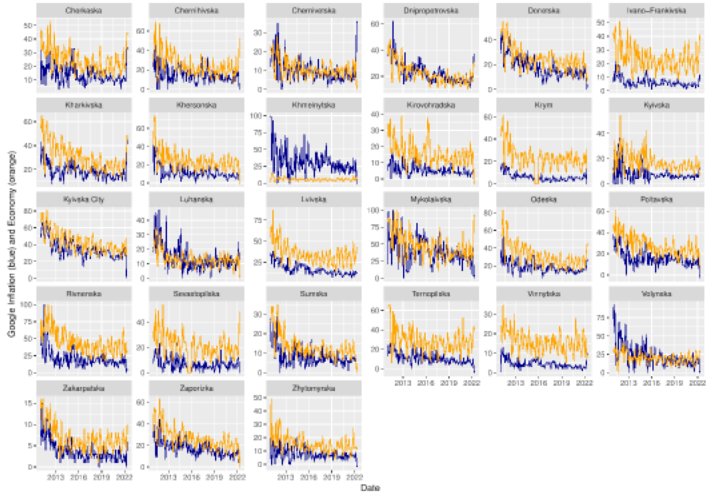
https://github.com/dankaufmann/f-curve

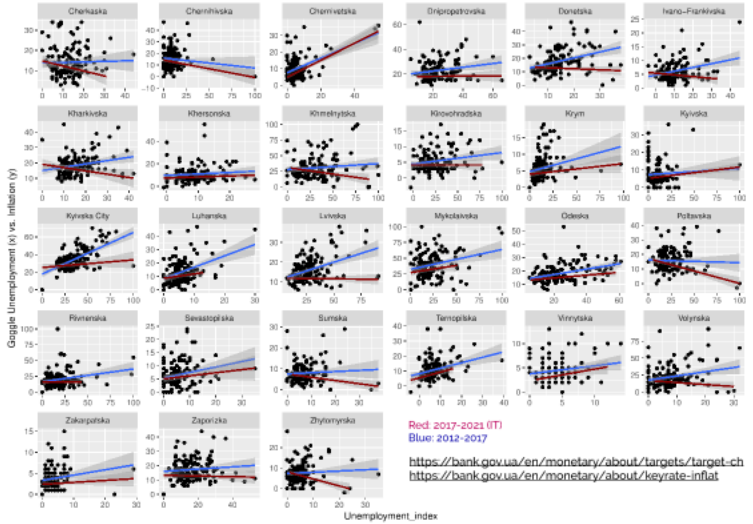
$$[X\ X^*] = [f\ f^*] \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & \lambda_{22} \end{bmatrix} + e$$

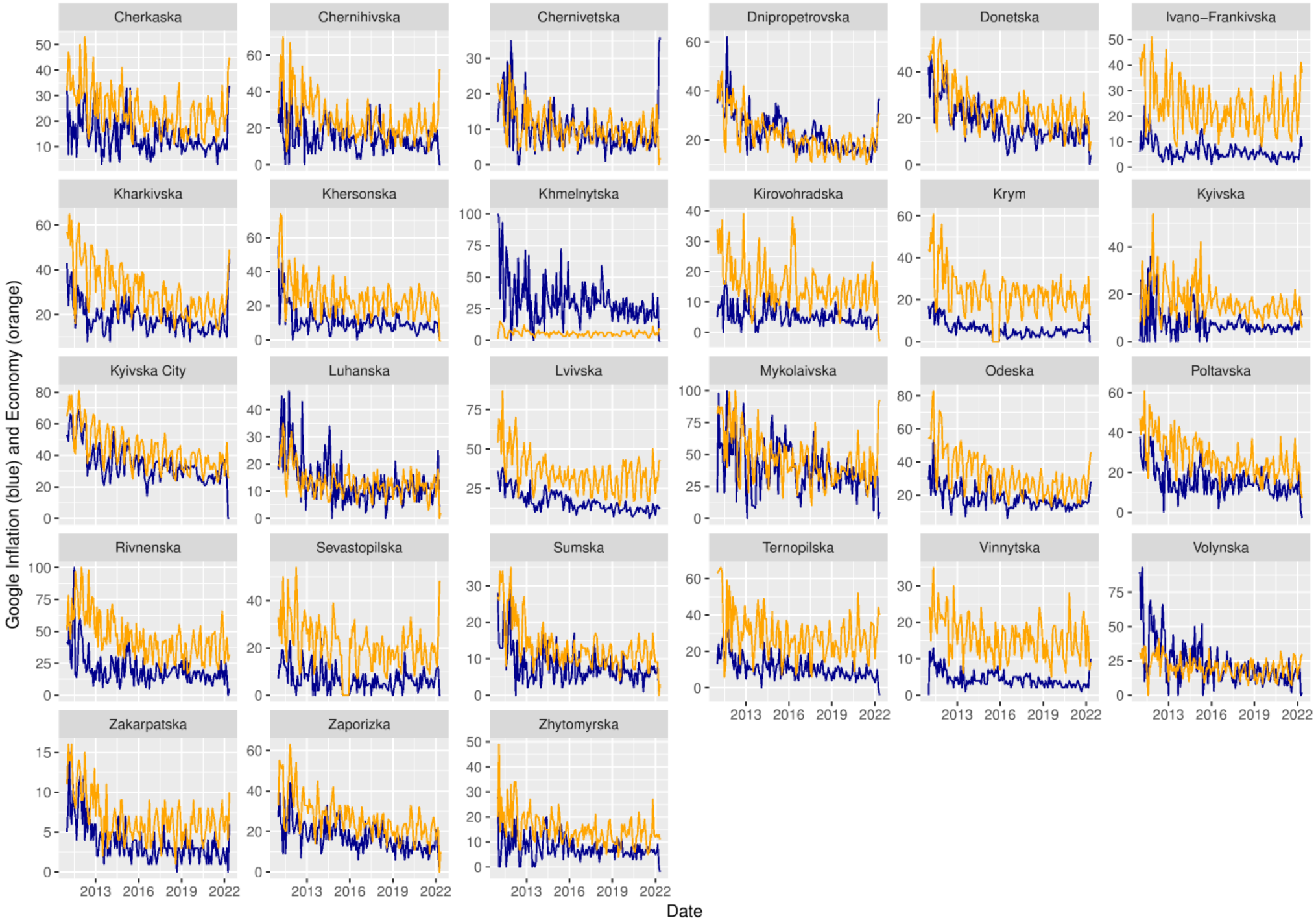
where X, X^* denote the data matrices comprising domestic and foreign variables, respectively. In addition f, f^* represent the domestic and foreign factors and $\lambda_{11}, \lambda_{21}, \lambda_{22}$ are the loading matrices.

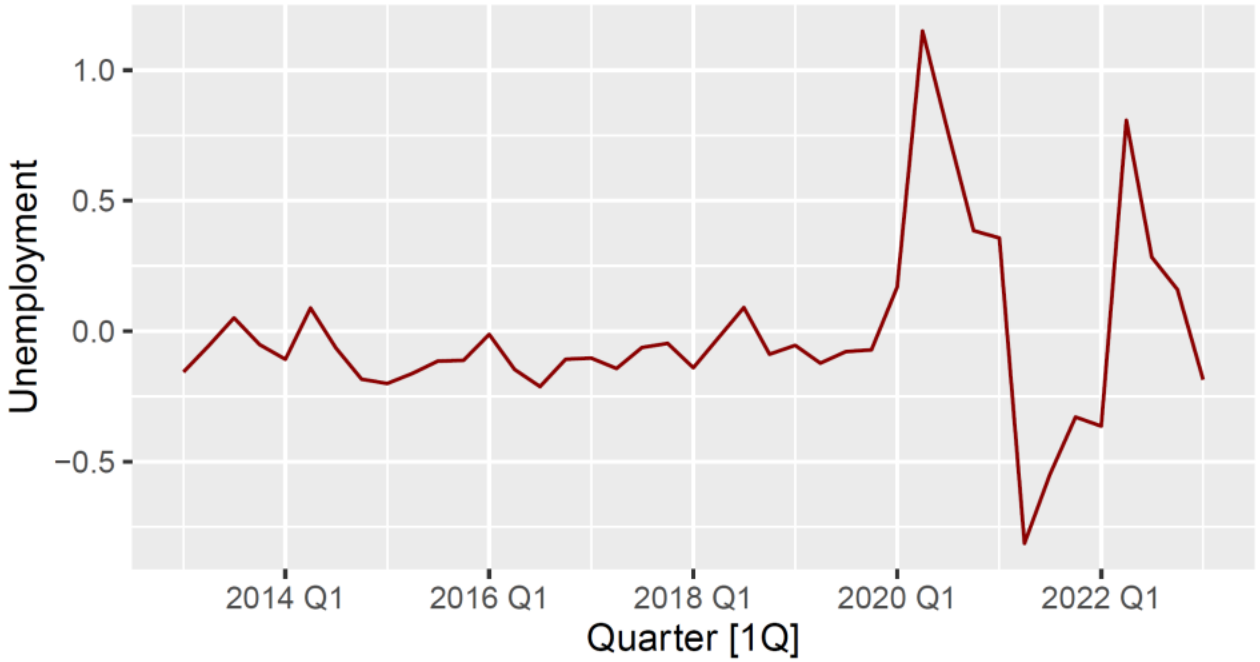
$$y_{\tau+h} = \alpha_h + \beta_{h,1} f_{\tau|t} + \beta_{h,2} f_{\tau-1} + v_{\tau+h}$$

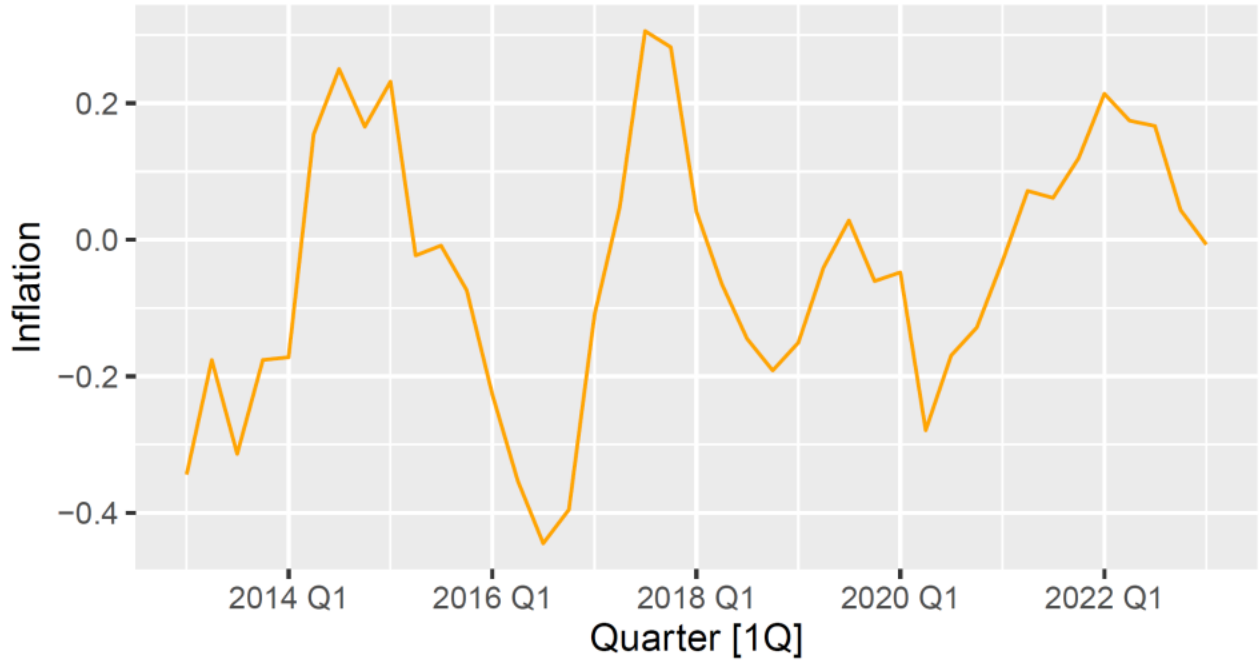
Google Trends - empirics

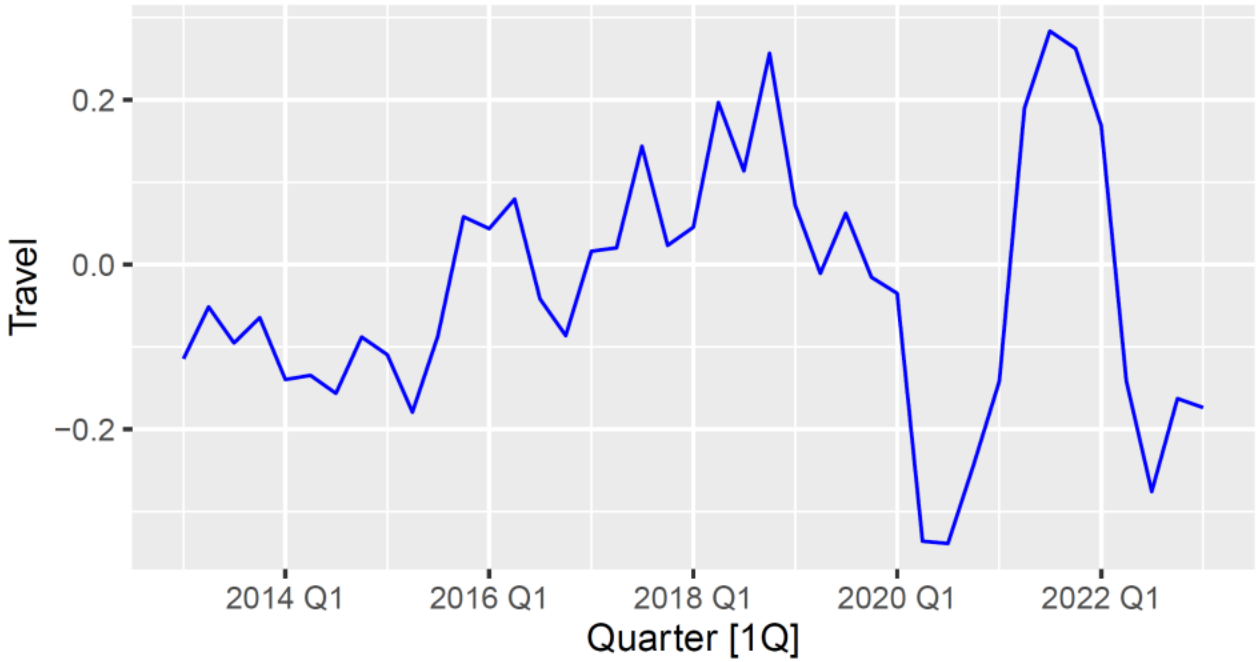


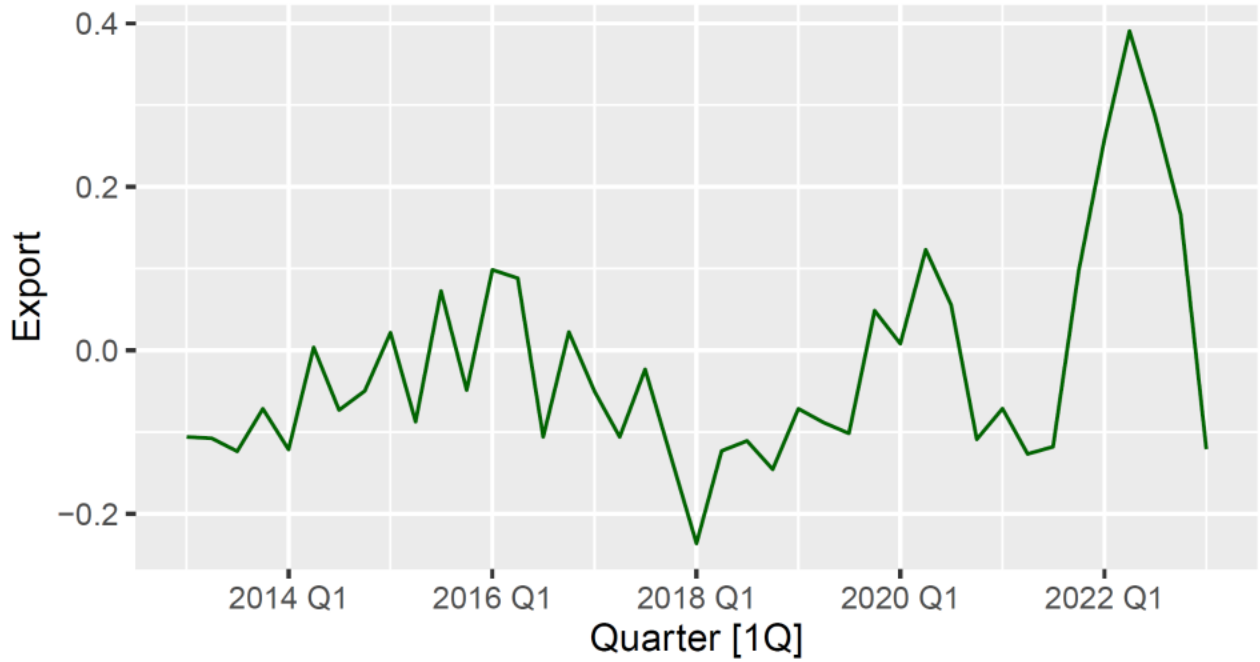


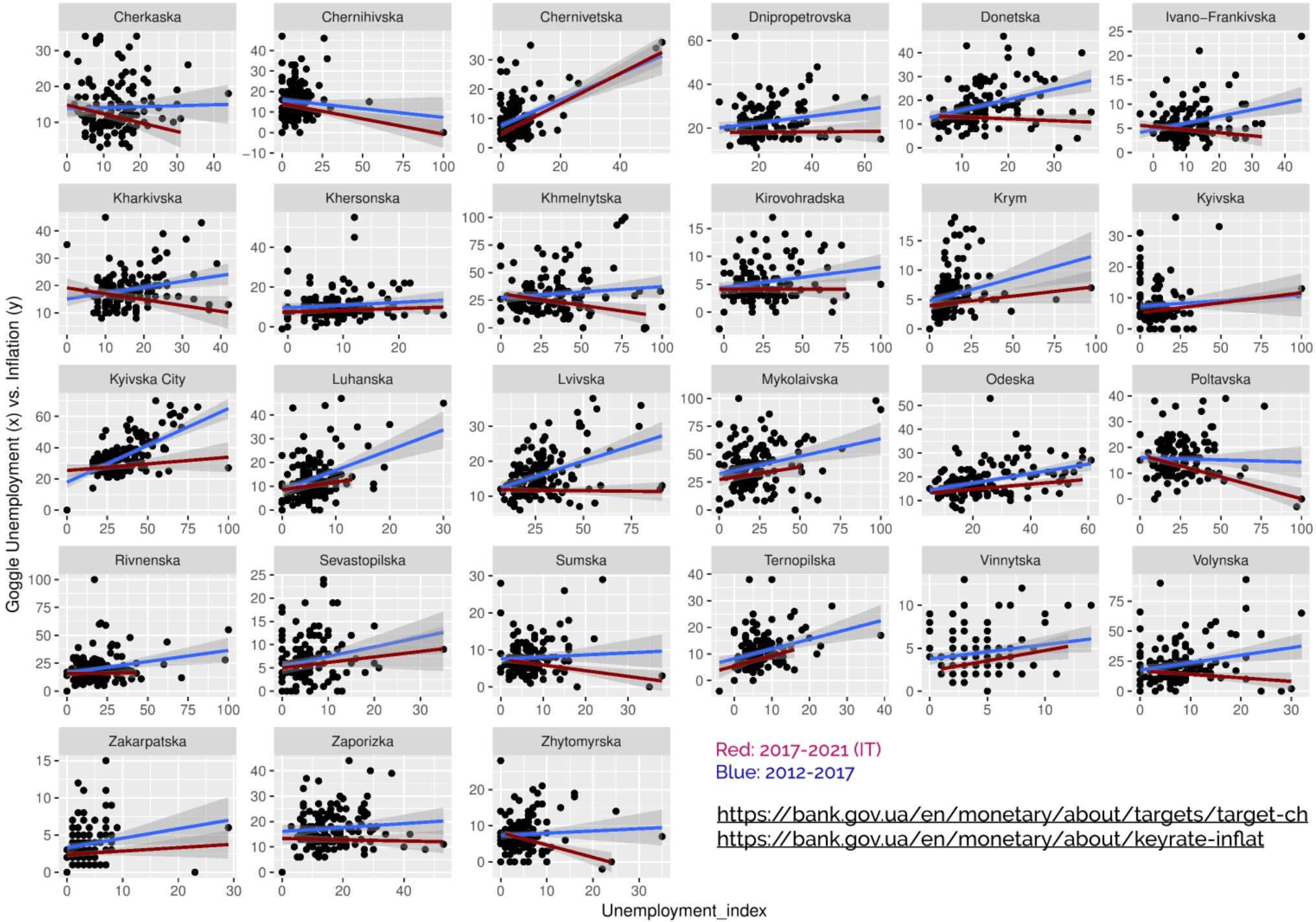

Cherkaska
Chernihivska
Chernivetska
Dnipropetrovska
Donetska
Ivano−Frankivska
Kharkivska
Khersonska
Khmelnytska
Kirovohradska
Krym
Kyivska
Kyivska City
Luhanska
Lvivska
Mykolaivska
Odeska
Poltavska
Rivnenska
Sevastopilska
Sumska
Ternopilska
Vinnytska
Volynska
Zakarpatska
Zaporizka
Zhytomyrska
Google Inflation (blue) and Economy (orange)
Date
2013
2016
2019
2022


Unemployment
1.0
0.5
0.0
−0.5
2014 Q1
2016 Q1
2018 Q1
2020 Q1
2022 Q1
Quarter [1Q]


Inflation
0.2
0.0
−0.2
−0.4
2014 Q1
2016 Q1
2018 Q1
2020 Q1
2022 Q1
Quarter [1Q]

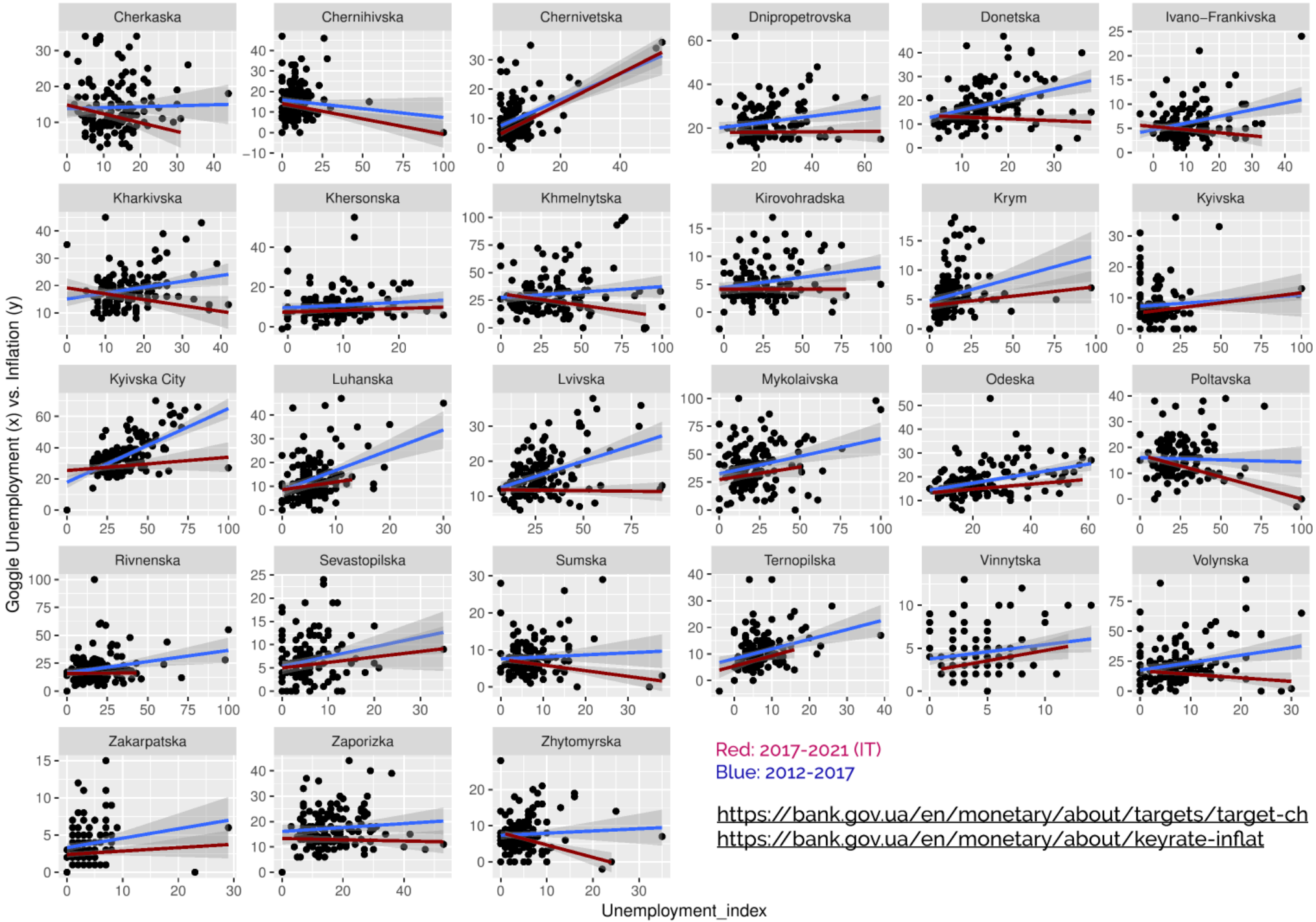

Travel
0.2
0.0
−0.2
2014 Q1
2016 Q1
2018 Q1
2020 Q1
2022 Q1
Quarter [1Q]


Export
0.4
0.2
0.0
−0.2
2014 Q1
2016 Q1
2018 Q1
2020 Q1
2022 Q1
Quarter [1Q]



Red: 2017-2021 (IT)
Blue: 2012-2017

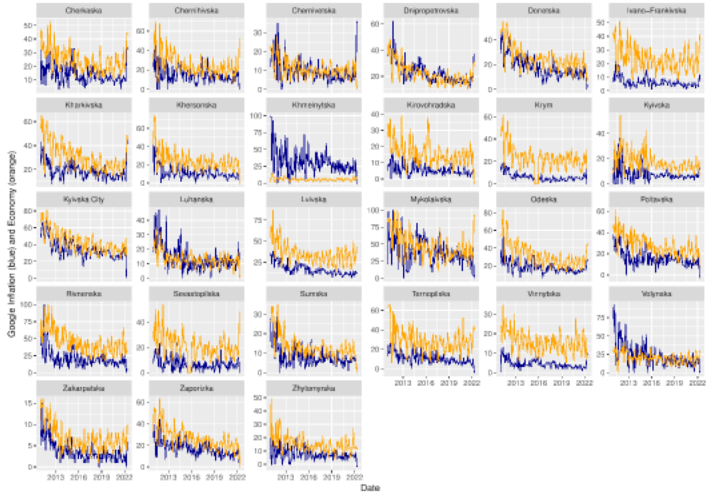
https://bank.gov.ua/en/monetary/about/targets/target-ch
https://bank.gov.ua/en/monetary/about/keyrate-inflat



Red: 2017-2021 (IT)
Blue: 2012-2017

https://bank.gov.ua/en/monetary/about/targets/target-ch
https://bank.gov.ua/en/monetary/about/keyrate-inflat

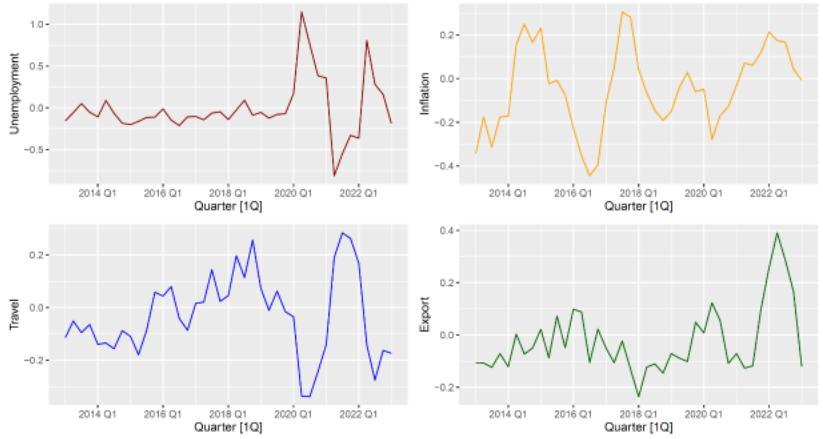
Google Trends - empirics

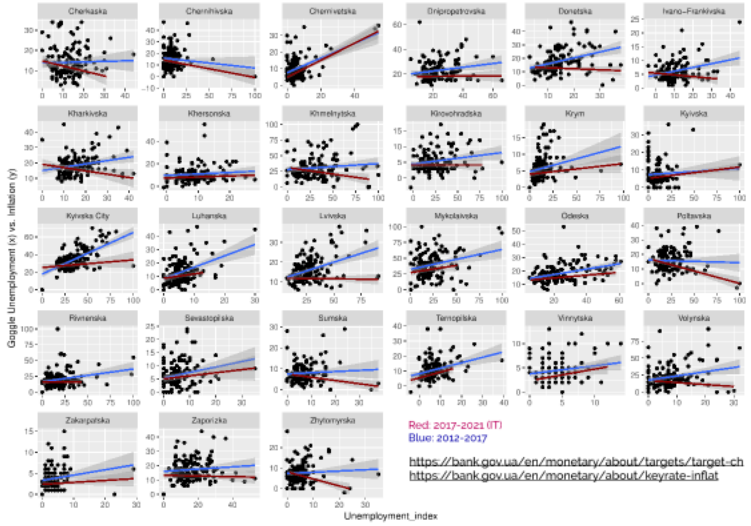


-> About 35 categories are used

-> Monthly time series contain plenty of variation (too much?) >> Q

-> Cross-index correlations appear reasonable at first sight but large number implies some shrinkage/ dimensionality reduction is needed



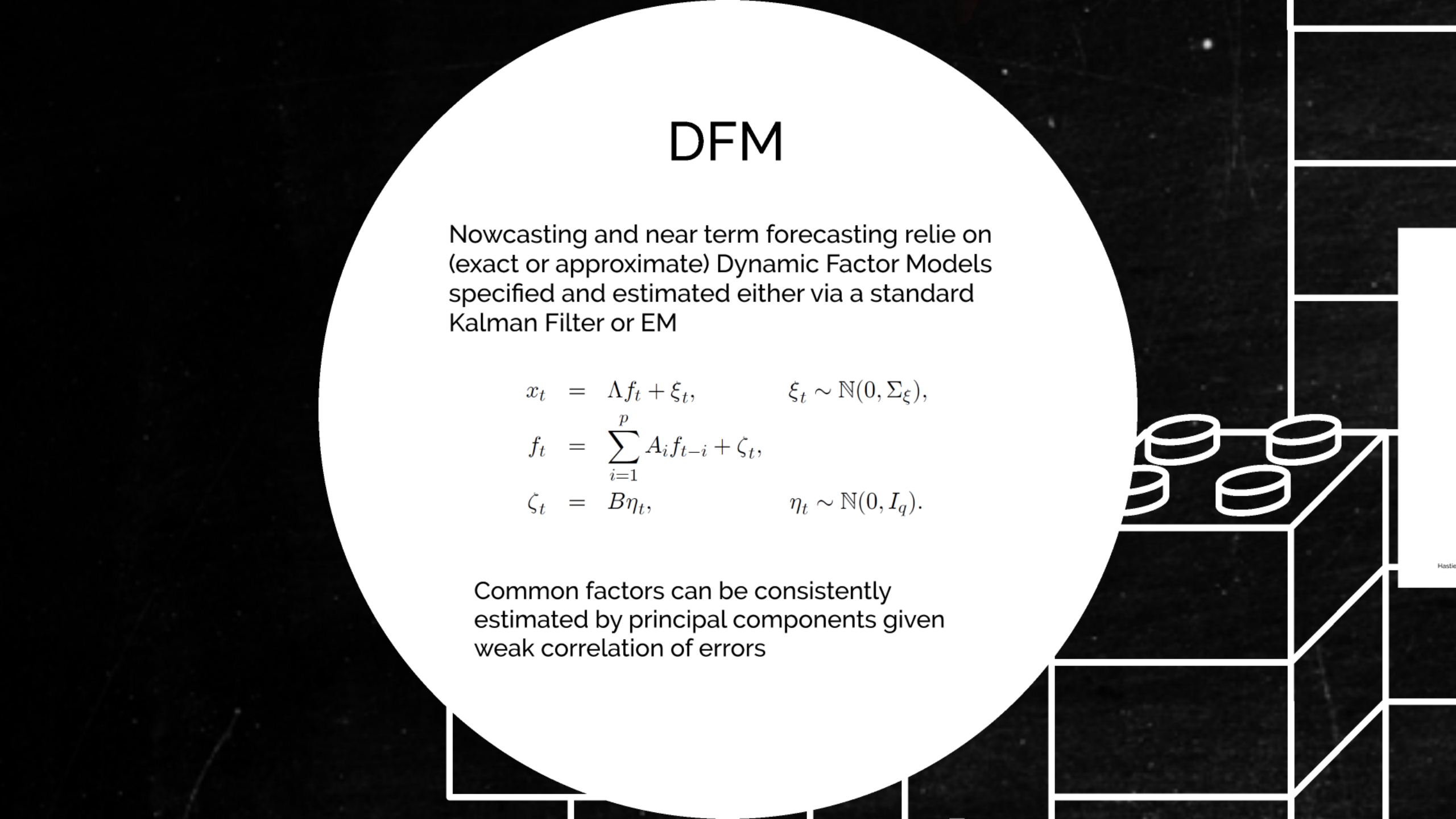


DFM

Nowcasting and near term forecasting relie on (exact or approximate) Dynamic Factor Models specified and estimated either via a standard Kalman Filter or EM

$$
\begin{aligned}
x_t &= \Lambda f_t + \xi_t, & \xi_t &\sim \mathbb{N}(0, \Sigma_\xi), \\
f_t &= \sum_{i=1}^{p} A_i f_{t-i} + \zeta_t, & & \\
\zeta_t &= B\eta_t, & \eta_t &\sim \mathbb{N}(0, I_q).
\end{aligned}
$$

Common factors can be consistently estimated by principal components given weak correlation of errors

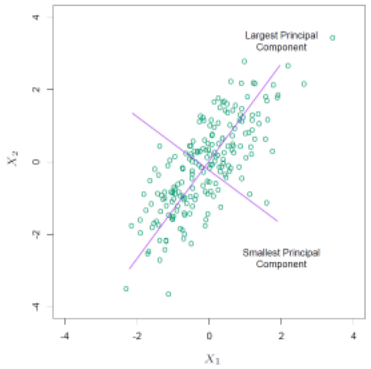


PCR - digging deeper

Principal Components

-> Given a data matrix **X** (N obs x p variables). PCA will perform a SVD of the centered matrix **X*** to find directions in the column space of **X*** that have small variance (with direction vectors *v* independent of each other)

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



-> In a subsequent stage, the PCs are used as inputs in a regression model, state-space model, etc

Hastie, Tibshirani, Friedman 2nd ed. "The Elements of Statistical Learning"

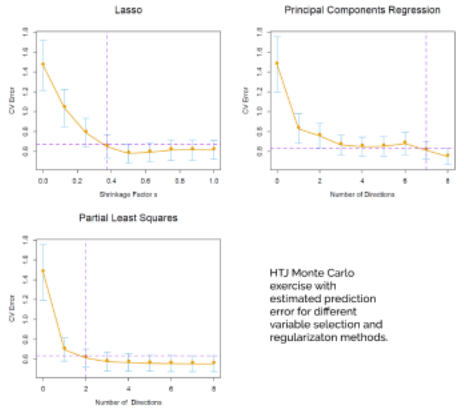
Partial Least Squares

-> PLS is a supervised method which identifies the components or factors (phi) to be independent of each other but also have high correlation with a target **y**
{Wold et al. 1984}

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

Hastie, Tibshirani, Friedman 2nd ed. "The Elements of Statistical Learning"

PLS as a Latent Factor Model

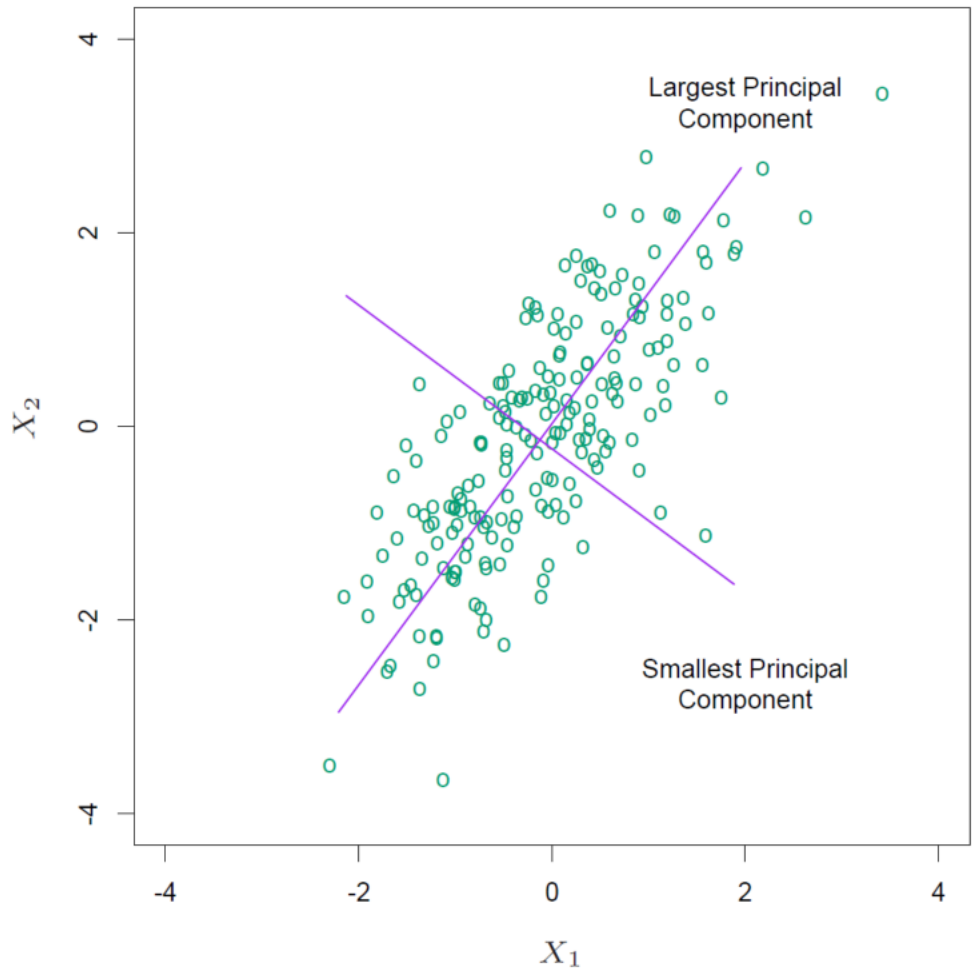


HTJ Monte Carlo exercise with estimated prediction error for different variable selection and regularizaton methods.

Principal Components

-> Given a data matrix **X** (N obs x p variables). PCA will perform a SVD of the centered matrix **X*** to find directions in the column space of **X*** that have small variance (with direction vectors *v* independent of each other)

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



-> In a subsequent stage, the PCs are used as inputs in a regression model, state-space model, etc

Partial Least Squares

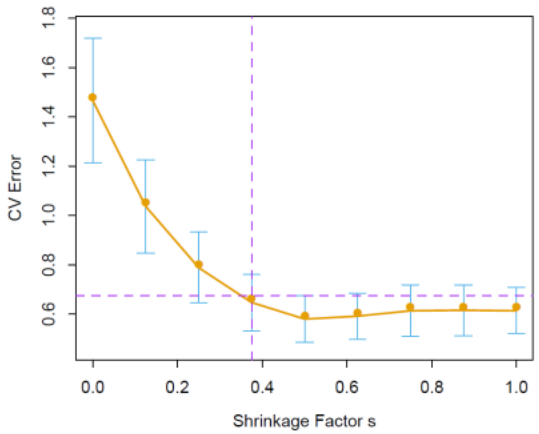
-> PLS is a supervised method which identifies the components or factors (phi) to be independent of each other but also have high correlation with a target **y**
{Wold et al. 1984}

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

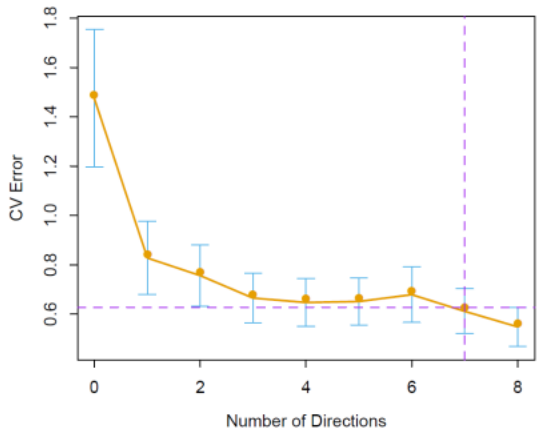
Hastie, Tibshirani, Friedman 2nd ed. "The Elements of Statistical Learning"

PLS as a Latent Factor Model

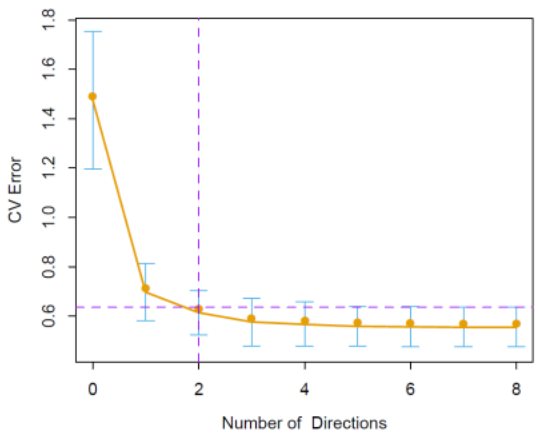
Lasso



Principal Components Regression



Partial Least Squares



HTJ Monte Carlo exercise with estimated prediction error for different variable selection and regularizaton methods.

PLS as a Latent Factor Model

PLS as a Latent Factor Model

$$
\begin{aligned}
T &= X \times W^*, \quad X \in \mathbb{R}^{n \times p}, T \in \mathbb{R}^{n \times K}, K < p \\
X &= T \times P' + \epsilon, \quad P \in \mathbb{R}^{p \times K} \\
y &= T \times C' + \xi, \quad C \in \mathbb{R}^{1 \times K}
\end{aligned}
$$

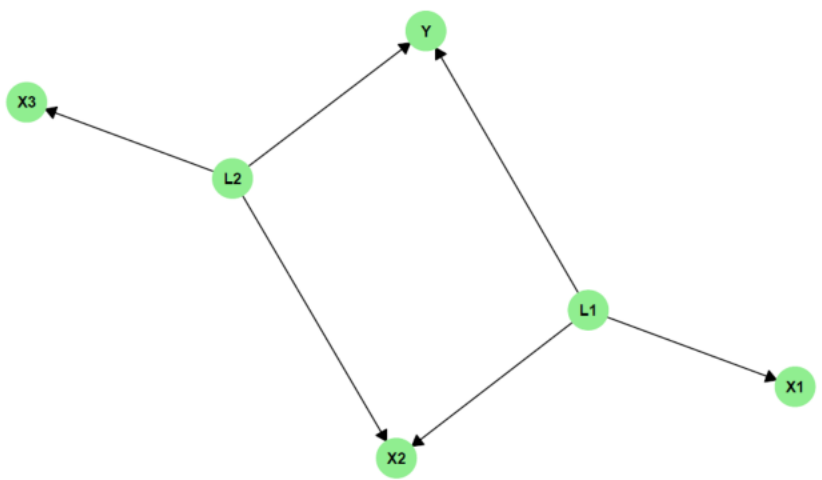


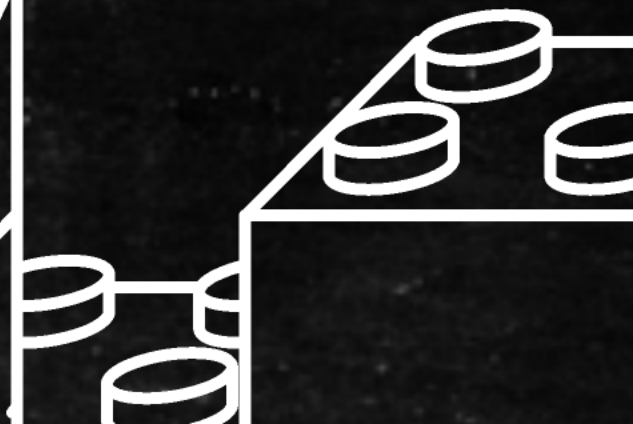
Figure 2 – Network Graph of a Simple Latent Factor Model

Modeling Avenues

Model is the result in equal parts of deliberate choice & data constraints

-> rich cross-section of potentially explanatory variables but too rich for a t=34 quarters for national GDP TS model; or a t=9 years, i=25 units panel (regional)

Considered:
-> Shrinkage: added benefit is interpretability of the ML model but unstable unless all Oblasts share the same parameters
-> PCR: reducing a large number of variables but loosing interpretability; lots of varations on the topic
-> PLS: collapsing large number of variables and improved fit, still loosing interpretability
- very large literature in bioinformatics, neuroscience on the use of PLS in "small n large p settings"
- usual genetics study has ~ 30x more variables than observations



Model is the result in equal parts of deliberate choice & data constraints

-> rich cross-section of potentially explanatory variables but too rich for a t=34 quarters for national GDP TS model; or a t=9 years, i=25 units panel (regional)

Considered:
-> Shrinkage: added benefit is interpretability of the ML model but unstable unless all Oblasts share the same parameters
-> PCR: reducing a large number of variables but loosing interpretability; lots of varations on the topic
-> PLS: collapsing large number of variables and improved fit, still loosing interpretability
- very large literature in bioinformatics, neuroscience on the use of PLS in "small n large p settings"
- usual genetics study has ~ 30x more variables than observations

ML model overview

Sparsity to tackle asymptotic inconsistency risk

-> Chun and Keles (2010) indicate challenges to asymptotic consistency of the PLS estimator in a "large p small n" context, with *fixed* **p1** *relevant* and *increasing* **p – p1** *irrelevant* variables.

-> The intuition for the lack of asymptotic consistency comes from the ridge-like nature of the PLS algorithm. Given that PLS latent factors load on all variables available in X, a larger fraction of irrelevant variables weaken the ability of the algorithm to identify the true factor directions.

-> Sparsity is achieved via variable selection in a multitude of ways, depending on the joint specificities of data sample and machine learning model (Lasso like, FWD or BWD Variable Selection, GA)

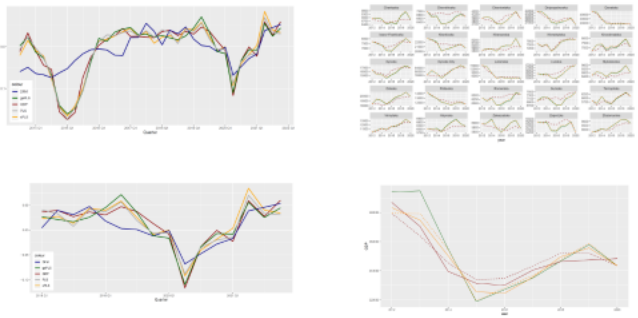
Variable Selection - overview

-> A wrapper (GA) and an embbeded method (sPLS) are used to induce sparsity: as a results, variable selection leads to improved interpretability

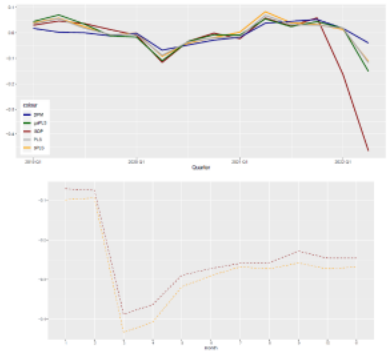
-> sPLS of Chun and Keles (2010) introduces a LASSO penalty int he optimization problem and jointly selects the optimal number of latent factors and the amount of penaty

Variable Selection - the GA algorithm

Insample Fit



Out of sample Forecast



-> Chun and Keles (2010) indicate challenges to asymptotic consistency of the PLS estimator in a "large p small n" context, with *fixed* **p1** *relevant* and *increasing* **p – p1** *irrelevant* variables.

-> The intuition for the lack of asymptotic consistency comes from the ridge-like nature of the PLS algorithm. Given that PLS latent factors load on all variables available in X, a larger fraction of irrelevant variables weaken the ability of the algorithm to identify the true factor directions.

-> Sparsity is achieved via variable selection in a multitude of ways, depending on the joint specificities of data sample and machine learning model (Lasso like, FWD or BWD Variable Selection, GA)

-> A wrapper (GA) and an embbeded method (sPLS) are used to induce sparsity: as a results, variable selection leads to improved interpretability

-> sPLS of Chun and Keles (2010) introduces a LASSO penalty int he optimization problem and jointly selects the optimal number of latent factors and the amount of penaty

Genes, chromosomes and Populations

time 0

	V1	V2	V3	V4	V5	*fit*
A1	0	0	0	0	0	*3*
A2	0	0	0	0	1	*12*
A...	...	...	...	...	...	
A32	1	1	1	1	1	*35*

time n

	V1	V2	V3	V4	V5	*fit*
A08	0	1	1	0	1	*31*
A13	0	1	0	1	1	*27*
A...	...	...	...	...	...	
A32	1	1	1	1	1	*35*

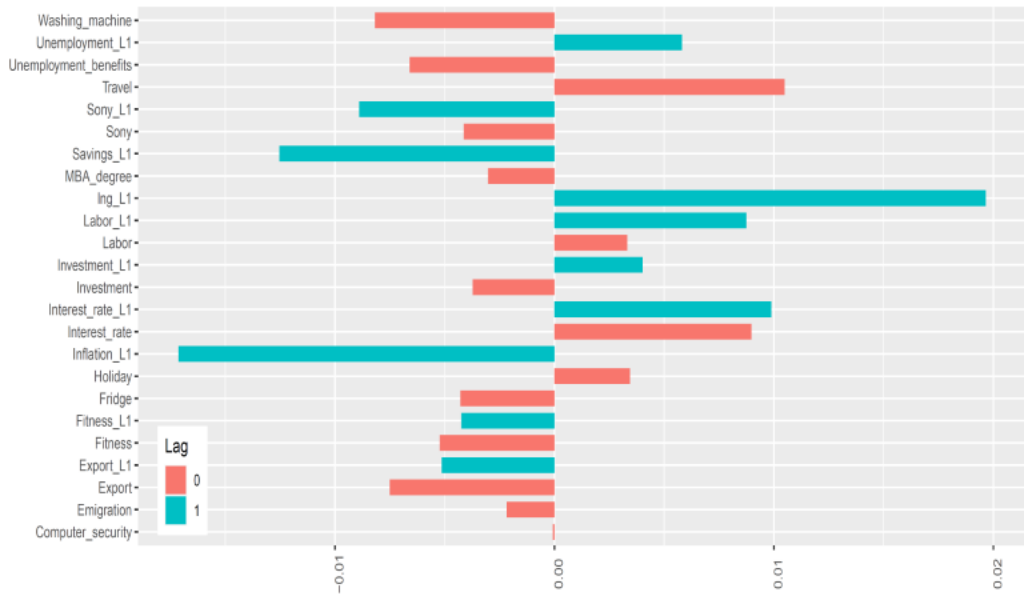
Goldberg "Genetic Algorithms in Search and Optimization"

P1	1	1	1	1	1
P1	0	0	0	0	0
C1	1	1	0	0	0
C2	0	0	1	1	1
C2'	0	1	0	1	1

Crossover: mixing of good genes from fit parents at random cross-over point (P1&P2 produce C1&C2)

&

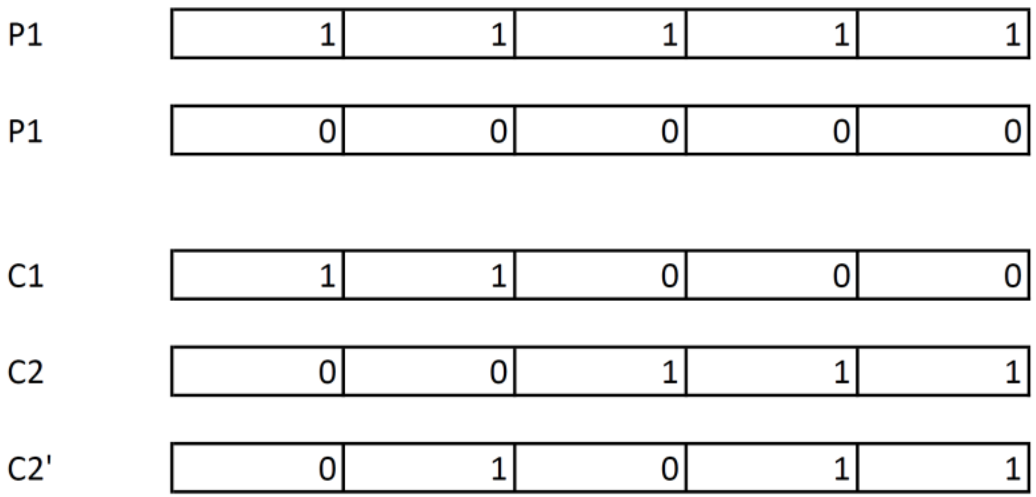
Mutation: random flips of certain genes (C2')



Genes, chromosomes and Populations

time 0

	V1	V2	V3	V4	V5	*fit*
A1	0	0	0	0	0	*3*
A2	0	0	0	0	1	*12*
A…	…	…	…	…	…	
A32	1	1	1	1	1	*35*

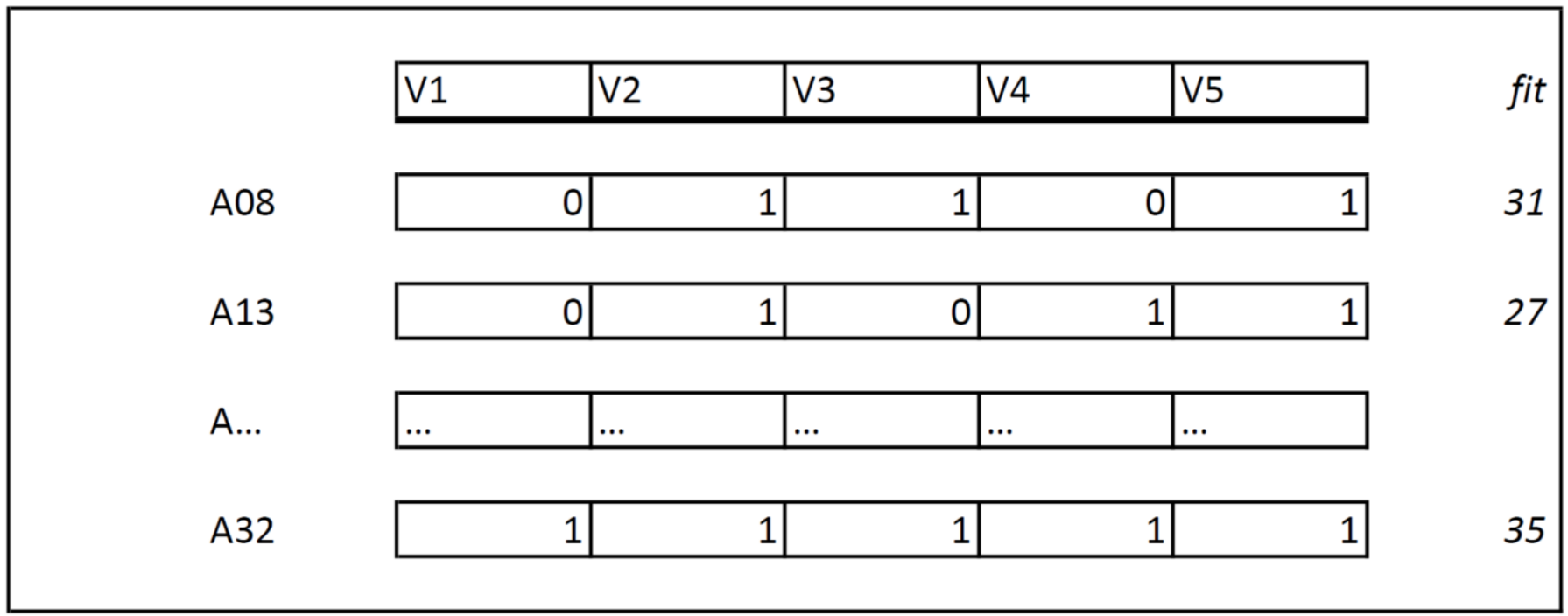


Crossover: mixing of good genes from fit parents at random cross-over point (P1&P2 produce C1&C2)

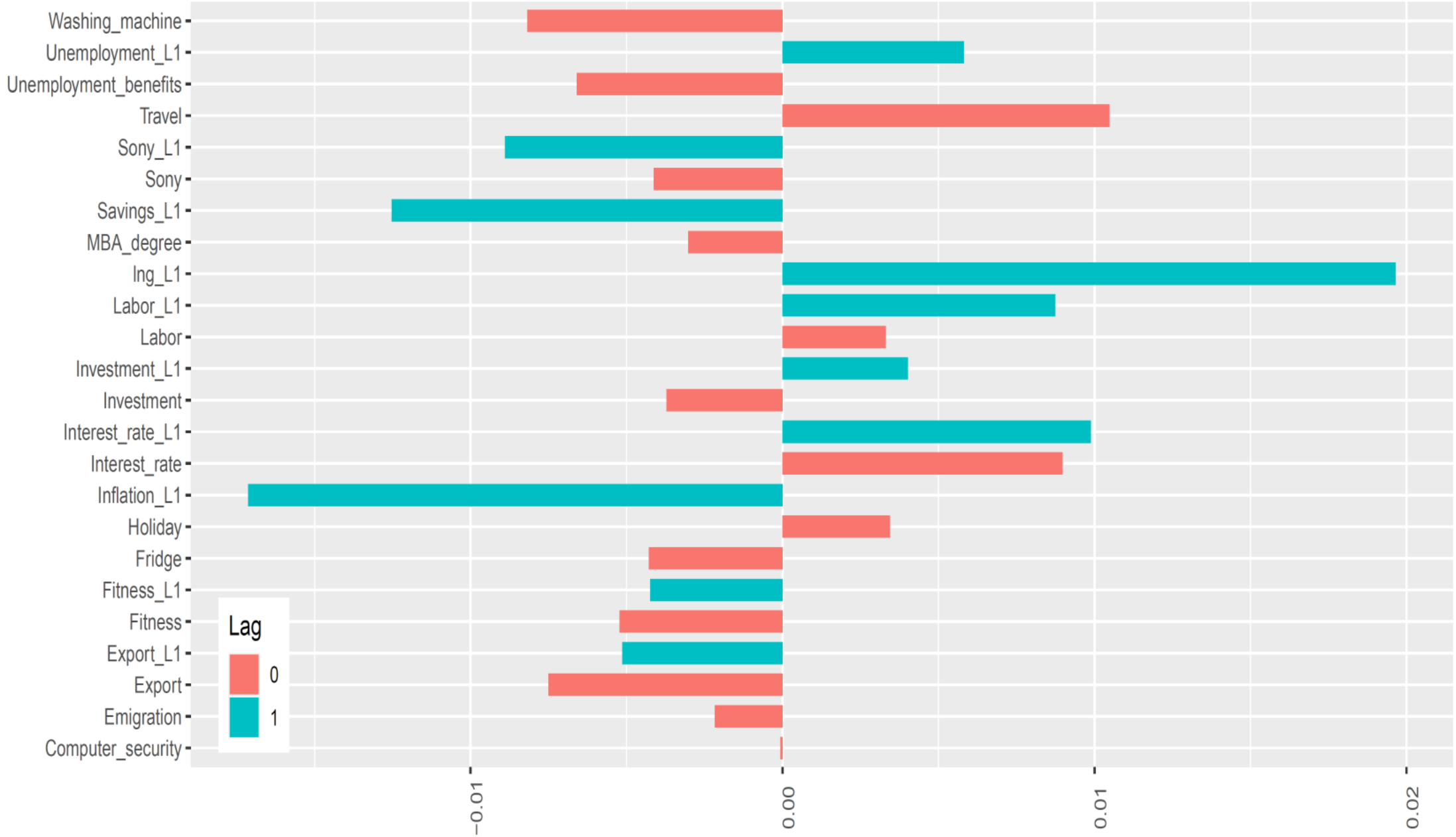
&

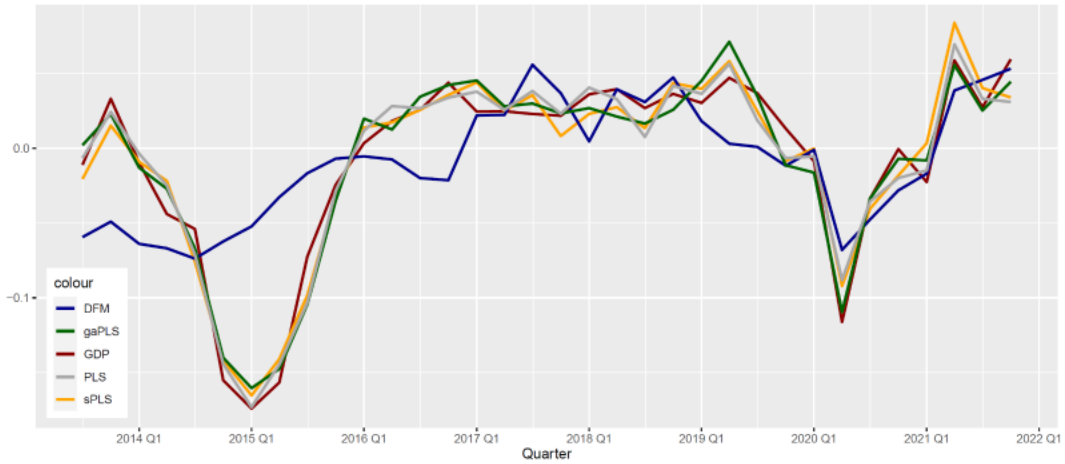
Mutation: random flips of certain genes (C2')

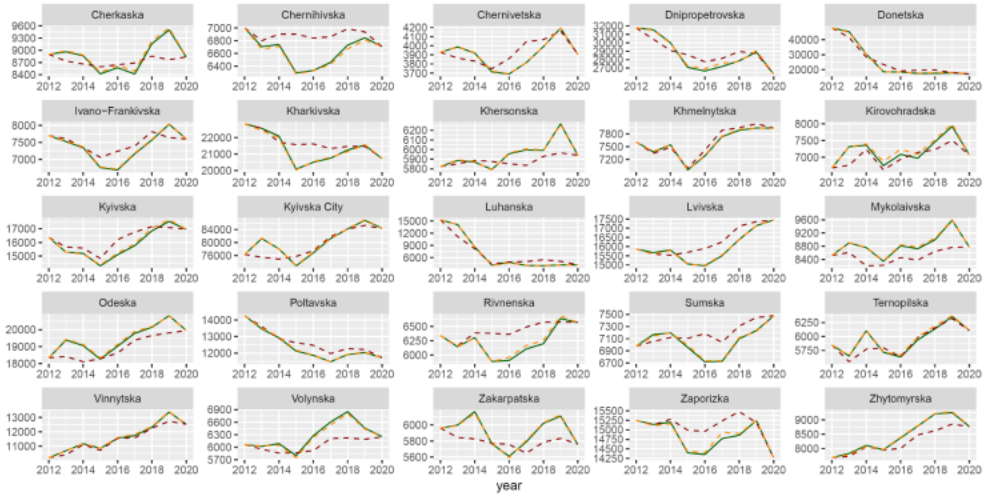
time n

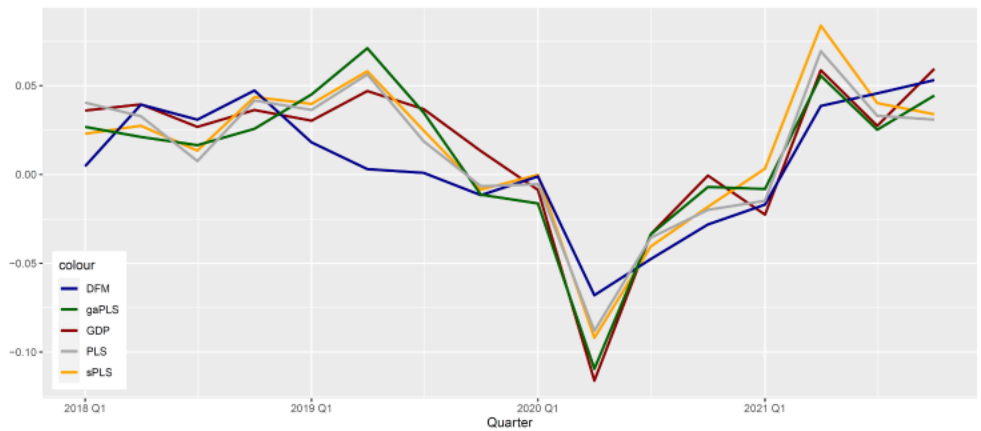


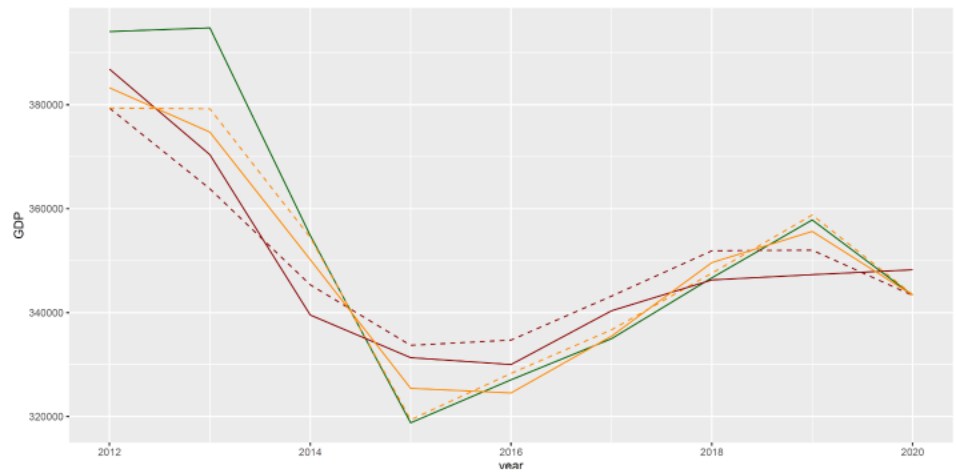
Goldberg "Genetic Algorithms in Search and Optimization"

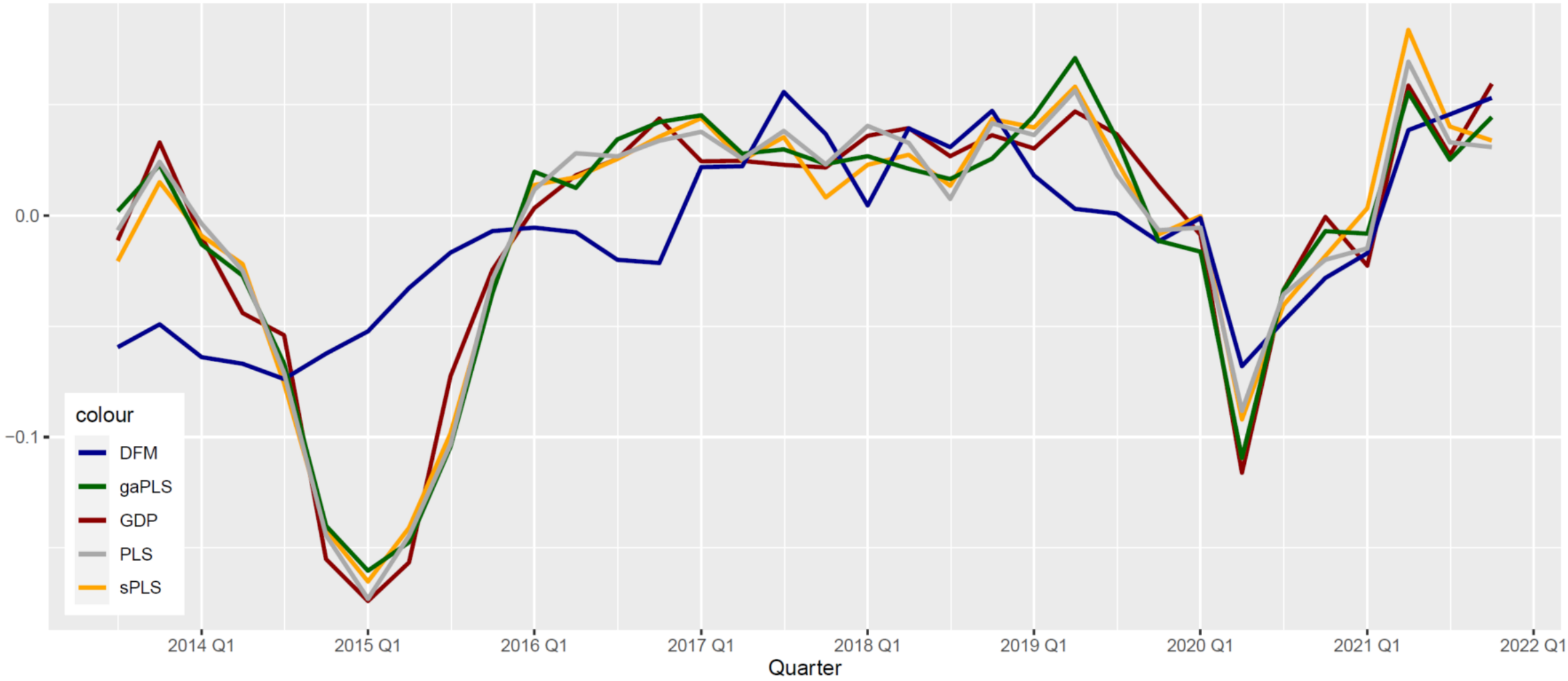

Washing_machine
Unemployment_L1
Unemployment_benefits
Travel
Sony_L1
Sony
Savings_L1
MBA_degree
Ing_L1
Labor_L1
Labor
Investment_L1
Investment
Interest_rate_L1
Interest_rate
Inflation_L1
Holiday
Fridge
Fitness_L1
Fitness
Export_L1
Export
Emigration
Computer_security
Lag
0
1
−0.01
0.00
0.01
0.02

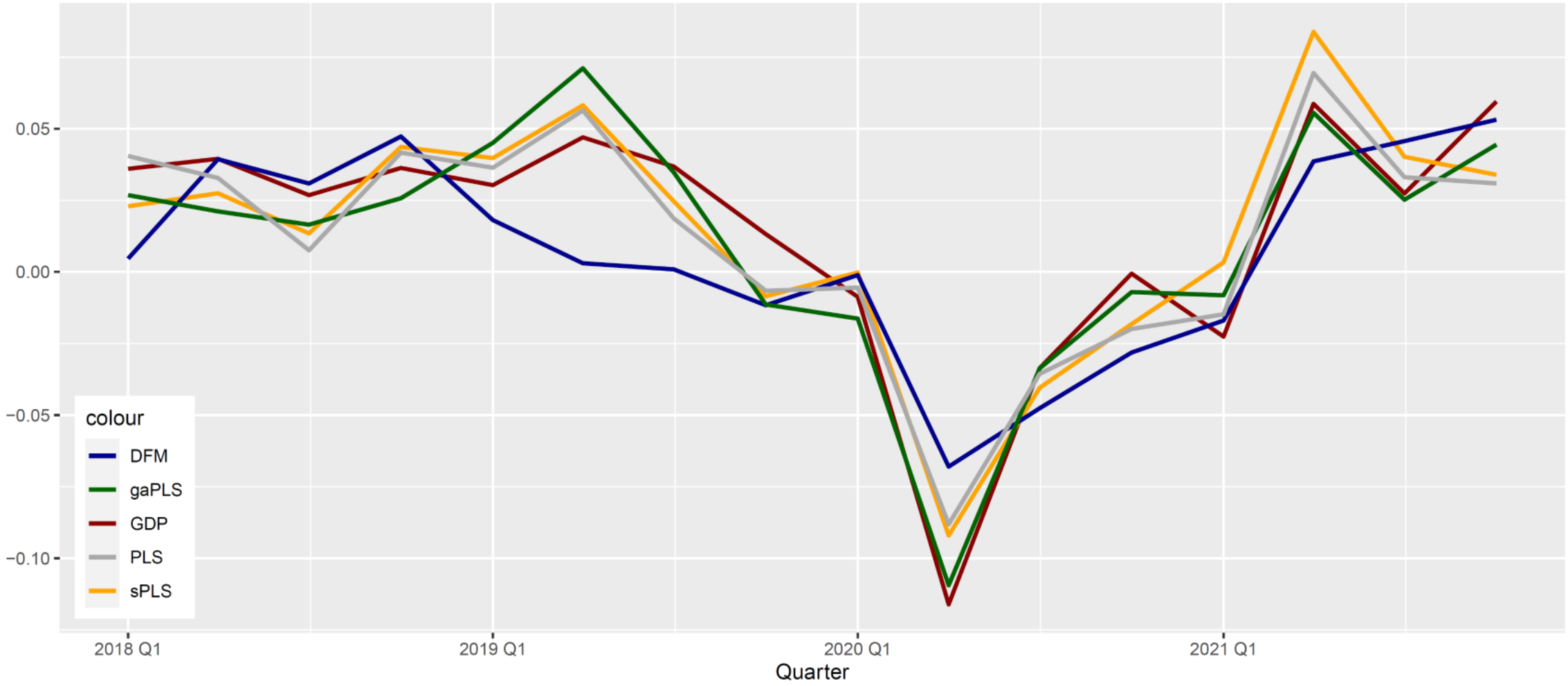

0.0
−0.1
colour
DFM
gaPLS
GDP
PLS
sPLS
2014 Q1
2015 Q1
2016 Q1
2017 Q1
2018 Q1
2019 Q1
2020 Q1
2021 Q1
2022 Q1
Quarter

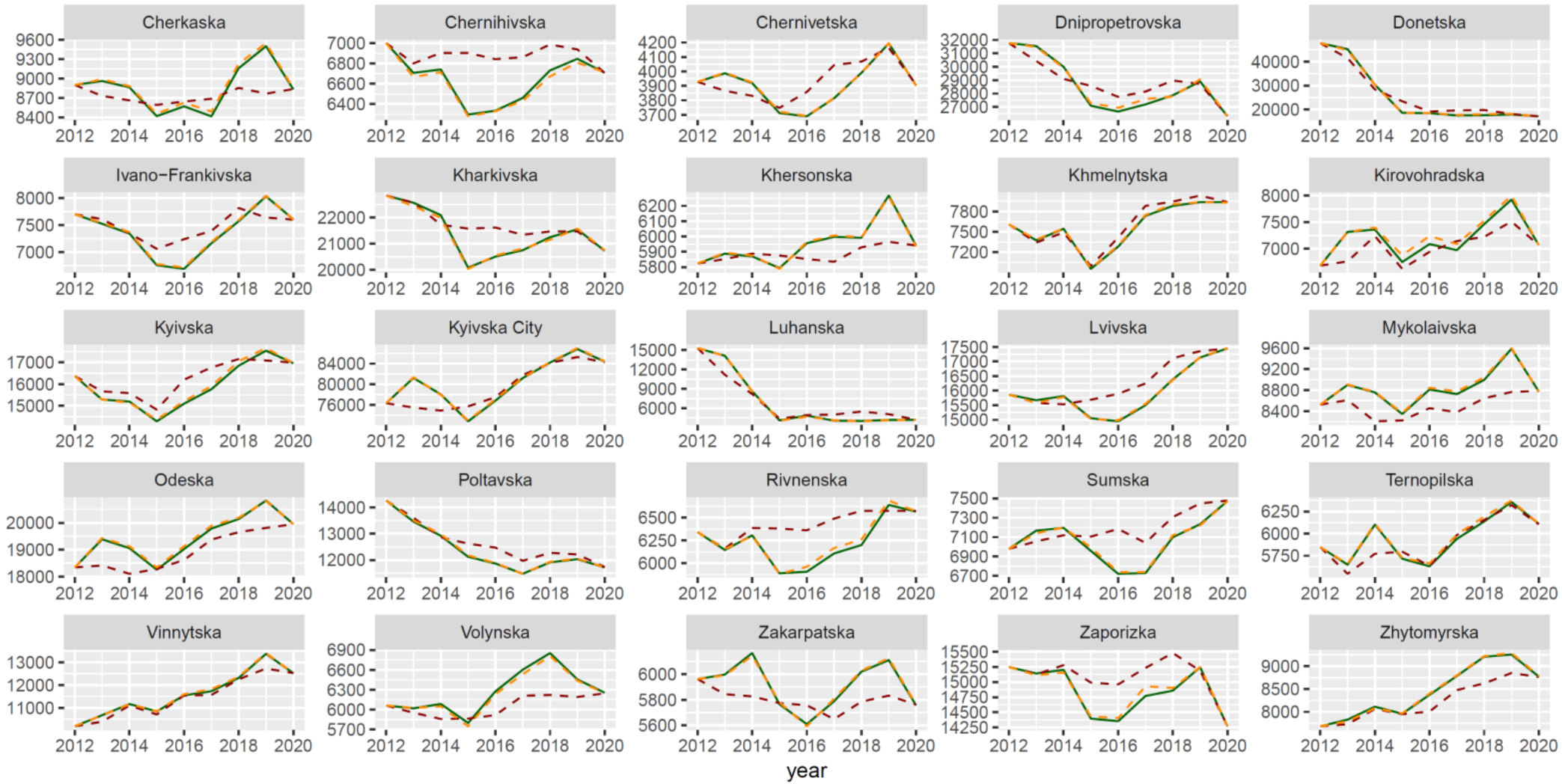

Cherkaska
Chernihivska
Chernivetska
Dnipropetrovska
Donetska
Ivano−Frankivska
Kharkivska
Khersonska
Khmelnytska
Kirovohradska
Kyivska
Kyivska City
Luhanska
Lvivska
Mykolaivska
Odeska
Poltavska
Rivnenska
Sumska
Ternopilska
Vinnytska
Volynska
Zakarpatska
Zaporizka
Zhytomyrska
2012
2014
2016
2018
2020
year

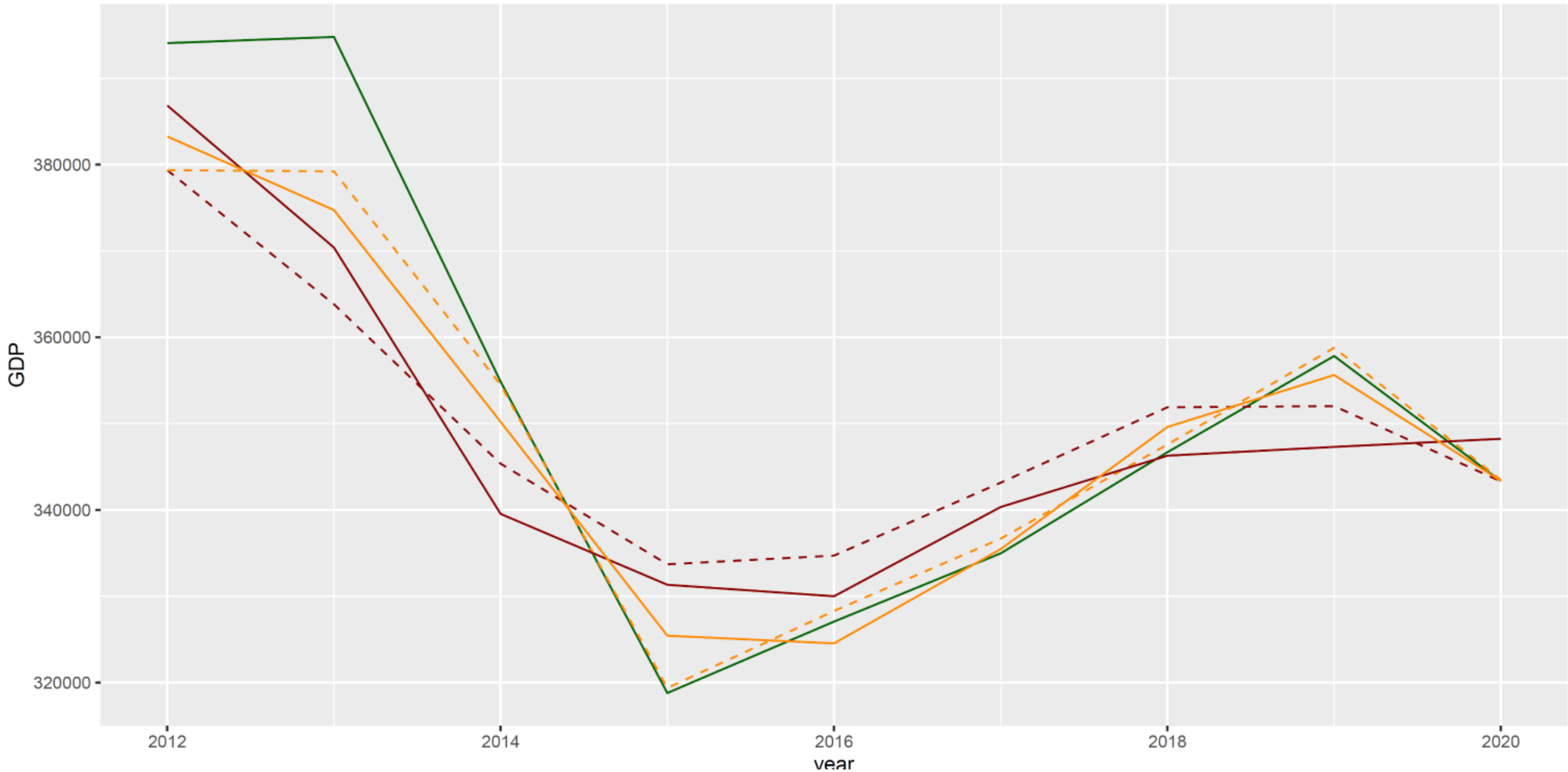

0.05
0.00
−0.05
−0.10
colour
DFM
gaPLS
GDP
PLS
sPLS
2018 Q1
2019 Q1
2020 Q1
2021 Q1
Quarter

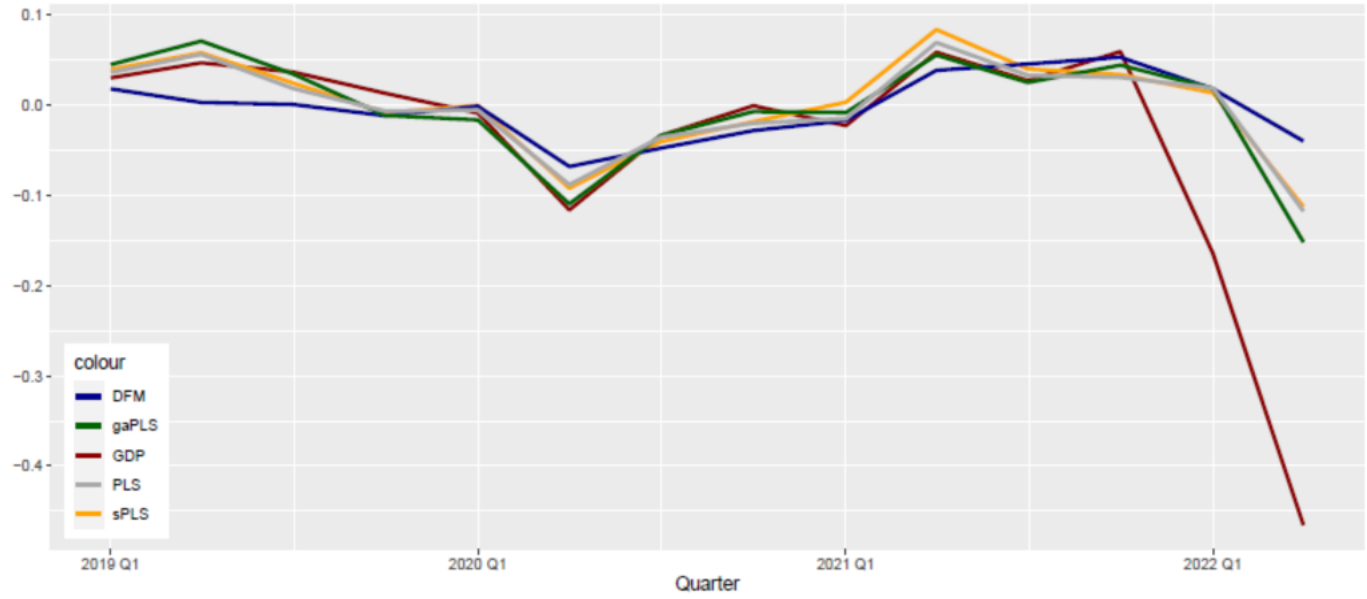

380000
360000
340000
320000
GDP
2012
2014
2016
2018
2020
year

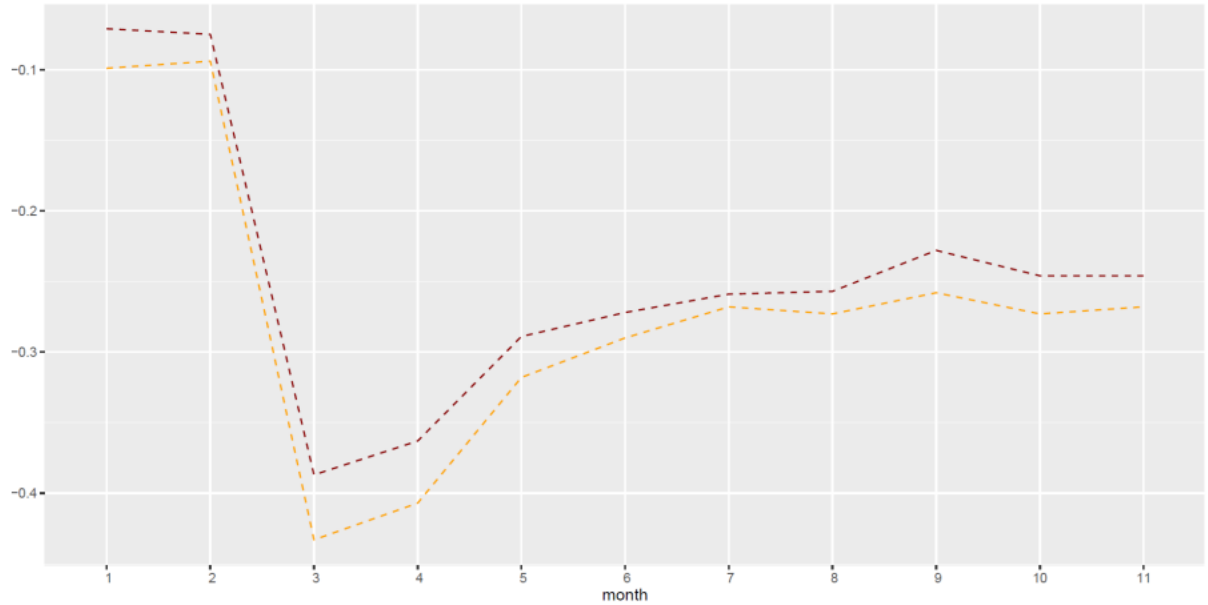

0.0
−0.1
colour
DFM
gaPLS
GDP
PLS
sPLS
2014 Q1
2015 Q1
2016 Q1
2017 Q1
2018 Q1
2019 Q1
2020 Q1
2021 Q1
2022 Q1
Quarter


0.05
0.00
−0.05
−0.10
colour
DFM
gaPLS
GDP
PLS
sPLS
2018 Q1
2019 Q1
2020 Q1
2021 Q1
Quarter


Cherkaska
Chernihivska
Chernivetska
Dnipropetrovska
Donetska
Ivano−Frankivska
Kharkivska
Khersonska
Khmelnytska
Kirovohradska
Kyivska
Kyivska City
Luhanska
Lvivska
Mykolaivska
Odeska
Poltavska
Rivnenska
Sumska
Ternopilska
Vinnytska
Volynska
Zakarpatska
Zaporizka
Zhytomyrska
year


GDP
380000
360000
340000
320000
2012
2014
2016
2018
2020
year


0.1
0.0
−0.1
−0.2
−0.3
−0.4
colour
DFM
gaPLS
GDP
PLS
sPLS
2019 Q1
2020 Q1
2021 Q1
2022 Q1
Quarter


−0.1
−0.2
−0.3
−0.4
1
2
3
4
5
6
7
8
9
10
11
month

final considerations

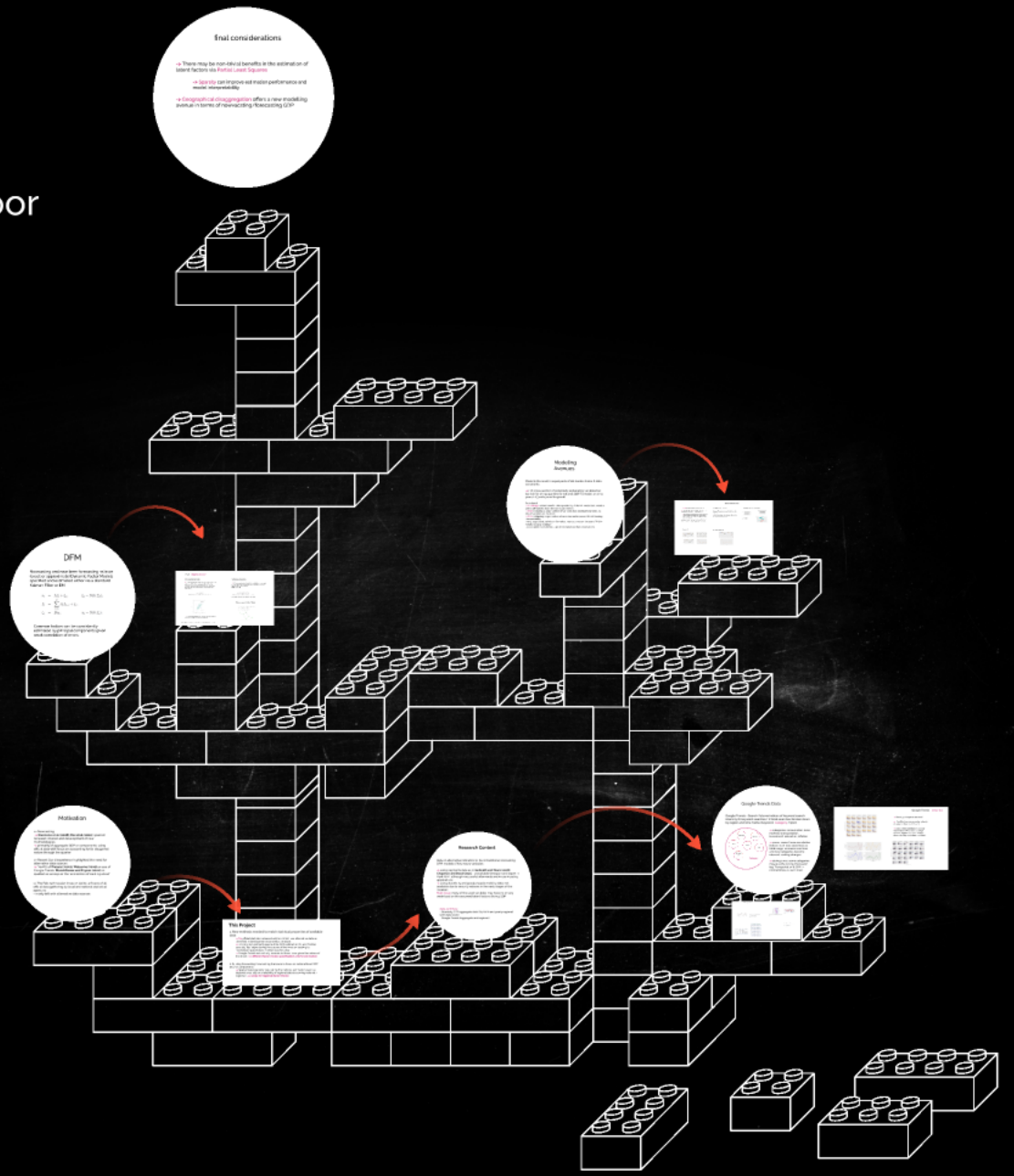
-> There may be non-trivial benefits in the estimation of latent factors via Partial Least Squares

-> Sparsity can improve estimation performance and model interpretability

-> Geographical disaggregation offers a new modelling avenue in terms of nowvacsting/forecasting GDP

Sparse Warcasting

Forecasting in a data-rich but statistics-poor environment



Mihnea Constantinescu
National Bank of Ukraine and
University of Amsterdam

June 2023