

Моделювання поточного прогнозування на базі змішано-частотного VAR підходу

ДОПОВІДАЧ: ЖУРАВЛЬОВА АНАСТАСІЯ

НАУКОВИЙ КЕРІВНИК: ДРІНЬ
СВІТЛАНА



Мета дослідження:

- Метою роботи є побудова моделі поточного прогнозування, яка інтегрує короткі дані у змішану частотну VAR модель для поточного прогнозування регіональних економічних показників України



Основні виклики при моделюванні:

1. **Проблема "рваного краю" (ragged edge):** Асинхронність надходження даних, коли на момент прогнозування t значення одних змінних вже відомі, тоді як інші ще не опубліковані.
2. **Обмежена довжина часових рядів:** Низка регіональних індикаторів має коротку історію спостережень, що ускладнює оцінку параметрів традиційних економетричних моделей та знижує їх прогностичну спроможність.

Використані дані для дослідження

:

Variable	Description	Frequency	Time Period	Release Lag	Source
income	Household income	Yearly	2001–2021	16M	UkrStat (2021d)
real_income	Disposable income	Yearly	2001–2021	16M	UkrStat (2021b)
salary_1	Wages	Yearly	2001–2020	14M	UkrStat (2021g)
sochelp	Social benefits	Yearly	2001–2020	16M	UkrStat (2021e)
expenses	Expenditures	Yearly	2001–2020	16M	UkrStat (2021c)
buying_goods	Retail goods purchases	Yearly	2001–2020	16M	UkrStat (2021e)
taxes	Income taxes	Yearly	2001–2020	16M	UkrStat (2021e)
socpay	Social contributions	Yearly	2001–2020	16M	UkrStat (2021e)
consume_expenses	NPISH consumption	Yearly	2008–2020	16M	UkrStat (2021e)
gross_kapital	Household capital formation	Yearly	2008–2020	16M	UkrStat (2021e)
metro_passangers	Subway passengers	Yearly	1995–2020	16M	UkrStat (2021f)
hired_workers	Hired employees	Yearly	2017–2018	16M	UkrStat (2018)
investments	Capital investments	Yearly	2019–2022	16M	UkrStat (2023)
realised_goods	Realized output	Yearly	2015–2016	10M	UkrStat (2016)
consumer_price_index	CPI (total)	Monthly	2001–2021	1M	MinFin (2021)
in_products	CPI: Food	Monthly	2020	1M	UkrStat (2020)
in_alco	CPI: Alcohol and tobacco	Monthly	2020	1M	UkrStat (2020)
in_clothes	CPI: Clothing	Monthly	2020	1M	UkrStat (2020)
in_apartments	CPI: Housing	Monthly	2020	1M	UkrStat (2020)
in_house_goods	CPI: Furnishings	Monthly	2020	1M	UkrStat (2020)
in_transport	CPI: Transport	Monthly	2020	1M	UkrStat (2020)
in_communication	CPI: Communication	Monthly	2020	1M	UkrStat (2020)
in_culture	CPI: Recreation and culture	Monthly	2020	1M	UkrStat (2020)
in_education	CPI: Education	Monthly	2020	1M	UkrStat (2020)
in_restaurants	CPI: Restaurants and hotels	Monthly	2020	1M	UkrStat (2020)
building	Construction output	Monthly	2012–2021	3M	UkrStat (2021a)
salary_2	Wages (monthly)	Monthly	2016	1M	UkrStat (2021g)
apple_gt	GT: "Apple"	Weekly	2020–2025	1D	Trends (2025)
economy_gt	GT: "Economy"	Weekly	2020–2025	1D	Trends (2025)
export_gt	GT: "Export"	Weekly	2020–2025	1D	Trends (2025)
gucci_gt	GT: "Gucci"	Weekly	2020–2025	1D	Trends (2025)
inflation_gt	GT: "Inflation"	Weekly	2020–2025	1D	Trends (2025)
investment_gt	GT: "Investment"	Weekly	2020–2025	1D	Trends (2025)
mercedes_gt	GT: "Mercedes"	Weekly	2020–2025	1D	Trends (2025)
saving_gt	GT: "Savings"	Weekly	2020–2025	1D	Trends (2025)
unemployment_gt	GT: "Unemployment"	Weekly	2020–2025	1D	Trends (2025)

Опис моделі MF-VAR:

$$\mathbf{y}_t = A_1 \mathbf{y}_{t-1} + A_2 \mathbf{y}_{t-2} + \dots + A_p \mathbf{y}_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma)$$

де:

- \mathbf{y}_t — вектор усіх спостережуваних змінних у системі на момент часу t ;
- $A_1, A_2, \dots, A_p \in \mathbb{R}^{K \times K}$ — матриці коефіцієнтів, які відображають, як минулі значення змінних впливають на поточні;
- $\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-p}$ — лагові значення змінних (наприклад, при $p = 2$, враховуються два попередні часові кроки);
- ε_t — випадковий шок або інновація, змодельований як білий шум: $\varepsilon_t \sim \mathcal{N}(0, \Sigma)$, де Σ — додатно визначена коваріаційна матриця;
- p — порядок лагу в системі.

Опис моделі MF-FAVAR:

Модель:

$$x_t^{(H)} = \Lambda^{(H)} f_t + \eta_t^{(H)}$$

$$y_t^{(L)} = \Lambda^{(L)} f_t + \eta_t^{(L)}$$

$$f_t = \Phi_1 f_{t-1} + \dots + \Phi_p f_{t-p} + \nu_t$$

Позначення:

$x_t^{(H)}$ — високочастотні змінні (CPI, Google Trends)

$y_t^{(L)}$ — низькочастотні змінні (GRP)

$f_t \in \mathbb{R}^r$ — латентні фактори

$\Lambda^{(H)}, \Lambda^{(L)}$ — матриці навантажень

Φ_i — коефіцієнти VAR, $\nu_t \sim \mathcal{N}(0, Q)$

Приклад (Київ):

$$x_t^{(H)} = [\text{Масго}_t, \text{GT}_t]^\top \in \mathbb{R}^2$$

$r = 2$, фактори з PCA або EMPCA

$$\text{Прогноз: } \hat{y}_t^{(L)} = \Lambda^{(L)} \hat{f}_t$$

Методи побудови факторів: PCA

Principal Components Analysis

Припустимо, ми маємо матрицю даних Z (розмірності $T \times k$), де T - кількість спостережень, k - кількість змінних. Матриця Z може містити пропущені значення.

Основна ідея факторного аналізу: представити Z як лінійну комбінацію латентних факторів F (розмірності $T \times r$, де $r \ll k$) та матриці факторних навантажень Λ (розмірності $k \times r$):

$$Z = \Lambda F' + E$$

де E - матриця помилок.

Методи побудови факторів: EMPCA

Expectation Maximization Principal Components Analysis

- E-крок (Expectation step):

$$\hat{Z}_{ij} = \sum_{s=1}^r \Lambda_{is}^{(l)} F_{sj}^{(l)}$$

- M-крок (Maximization step):

$$Z^{(l+1)} = U\Sigma V'$$

Методи побудови факторів: TW

Tall-Wide

- “Tall” block:

- Нехай Z_{Tall} - підматриця Z , що містить лише стовпці без пропущених значень.
- PCA на Z_{Tall} :

$$Z_{Tall} = \hat{\Lambda}^{Tall} \hat{F}^{Tall} + E_{Tall}$$

де $\hat{\Lambda}^{Tall}$ - оцінка матриці факторних навантажень, \hat{F}^{Tall} - оцінка матриці факторів, E_{Tall} - матриця помилок.

- “Wide” block:

- Нехай Z_{Wide} - підматриця Z , що містить лише рядки без пропущених значень.
- PCA на Z_{Wide} :

$$Z_{Wide} = \hat{\Lambda}^{Wide} \hat{F}^{Wide} + E_{Wide}$$

де $\hat{\Lambda}^{Wide}$ - оцінка матриці факторних навантажень, \hat{F}^{Wide} - оцінка матриці факторів, E_{Wide} - матриця помилок.

Методи побудови факторів: TP

Tall-Project

Заповнення пропущених значень:

- Для кожної змінної i , що не входить до "високого" блоку, будується регресія:

$$Z_i^{obs} = \hat{F}^{Tall} \beta_i + e_i$$

де Z_i^{obs} - вектор спостережуваних значень змінної i , β_i - вектор коефіцієнтів регресії, e_i - вектор помилок.

- Пропущені значення для змінної i заповнюються прогнозами з цієї регресії:

$$\hat{Z}_i^{miss} = \hat{F}^{Tall} \hat{\beta}_i$$

де \hat{Z}_i^{miss} - вектор заповнених значень змінної i , $\hat{\beta}_i$ - оцінка вектора β_i .

Оцінка алгоритму:

Формула для RMSFE:

$$\text{RMSFE} = \sqrt{\text{MSE}}$$

де MSE (середньоквадратична помилка) розраховується як:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2$$

У цих формулах:

- n – кількість спостережень або точок даних.
- P_i – прогнозне значення для i -го спостереження.
- O_i – фактичне (спостережуване) значення для i -го спостереження.

Оцінка алгоритму:

Формула CRPS:

$$\text{CRPS}(F, x_a) = \int_{-\infty}^{\infty} (F(x) - H(x - x_a))^2 dx$$

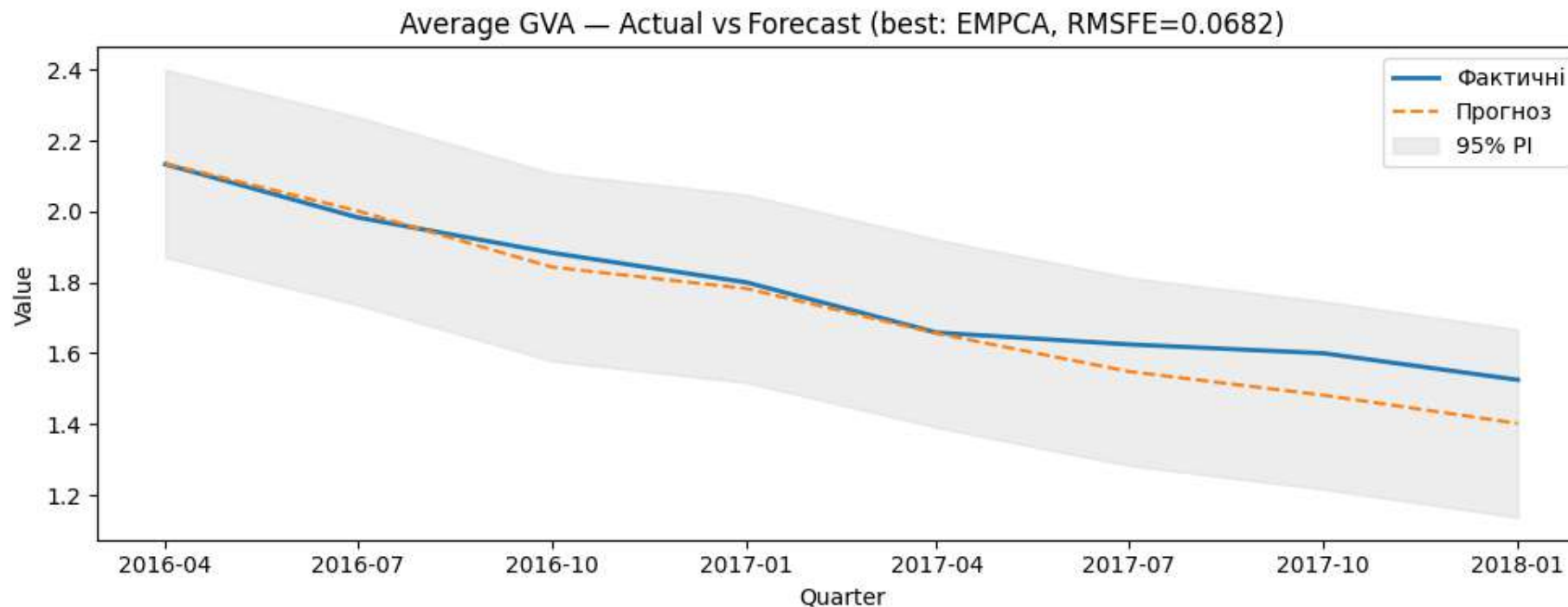
де:

- $F(x)$ – кумулятивна функція розподілу (CDF) прогнозу.
- x_a – фактичне спостережуване значення.
- $H(x - x_a)$ – функція Гевісайда, яка визначається як:

$$H(y) = \begin{cases} 0 & \text{якщо } y < 0 \\ 1 & \text{якщо } y \geq 0 \end{cases}$$

Отримані результати для моделі для Лондону:

EMPCA	: RMSFE = 0.1041	CRPS = 0.0625
TW	: RMSFE = 0.1194	CRPS = 0.0728
TP	: RMSFE = 0.1611	CRPS = 0.0941
BVAR	: RMSFE = 0.1601	CRPS = 0.7042

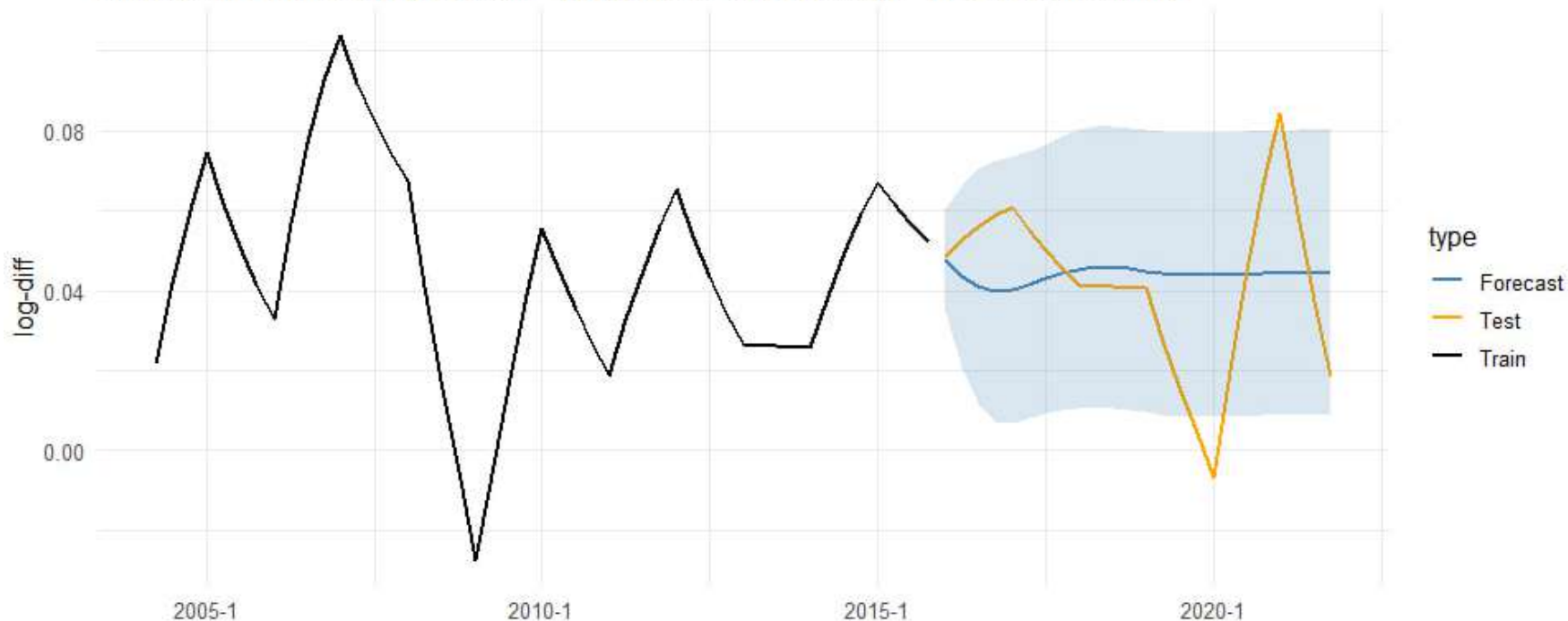


Отримані результати для моделі для Києва:

Kyiv quarterly GRP growth: train / test / best-MF-VAR forecast

Best method = EMPCA | Lag =2 | MAE=0.016 | RMSE=0.021 | SMAPE=43.51% | CRPS=0.012

Horizon	Method	Lag	MAE	RMSE	CRPS
24 qtrs	EMPCA	2	0.0158	0.0208	0.0116
	BPCA	2	0.0159	0.0208	0.0117
	SVDI	4	0.0255	0.0319	0.0194
4 qtrs	SVDI	4	0.0155	0.0167	0.0093
	BPCA	2	0.0211	0.0237	0.0134
	EMPCA	2	0.0221	0.0247	0.0141
2 qtrs	EMPCA	2	0.0033	0.0033	0.0034
	BPCA	2	0.0033	0.0034	0.0034
	SVDI	4	0.0053	0.0054	0.0042



Отримані результати для моделі для Києва:

Horizon	Method	Lag	MAE	RMSE	CRPS
24 qtrs	EMPCA	2	0.0158	0.0208	0.0116
	BPCA	2	0.0159	0.0208	0.0117
	SVDI	4	0.0255	0.0319	0.0194
4 qtrs	SVDI	4	0.0155	0.0167	0.0093
	BPCA	2	0.0211	0.0237	0.0134
	EMPCA	2	0.0221	0.0247	0.0141
2 qtrs	EMPCA	2	0.0033	0.0033	0.0034
	BPCA	2	0.0033	0.0034	0.0034
	SVDI	4	0.0053	0.0054	0.0042

Kyiv quarterly GRP growth: train / test / best-MF-VAR forecast

Best method = SVDI | Lag =4 | MAE=0.016 | RMSE=0.017 | SMAPE=36.7% | CRPS=0.009

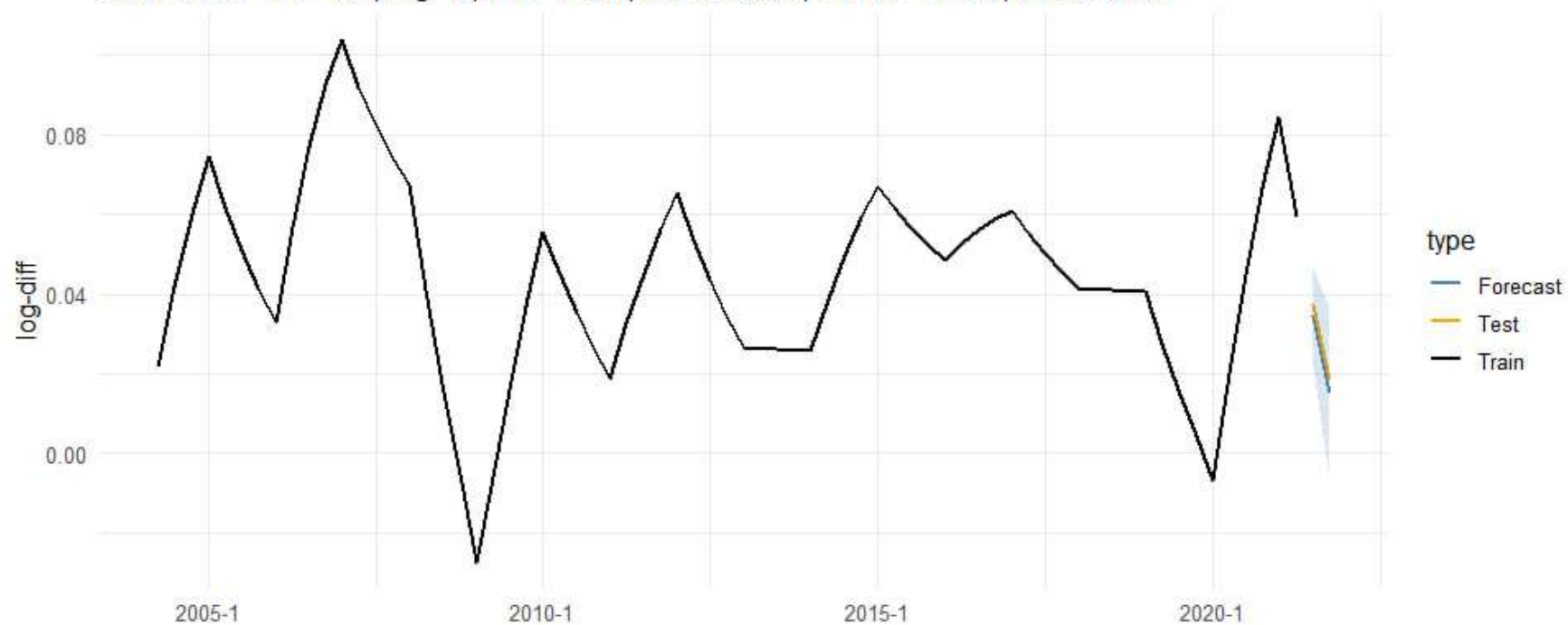


Отримані результати для моделі для Києва

Kyiv quarterly GRP growth: train / test / best-MF-VAR forecast

Best method = EMPCA | Lag =2 | MAE=0.003 | RMSE=0.003 | SMAPE=14.5% | CRPS=0.003

Horizon	Method	Lag	MAE	RMSE	CRPS
24 qtrs	EMPCA	2	0.0158	0.0208	0.0116
	BPCA	2	0.0159	0.0208	0.0117
	SVDI	4	0.0255	0.0319	0.0194
4 qtrs	SVDI	4	0.0155	0.0167	0.0093
	BPCA	2	0.0211	0.0237	0.0134
	EMPCA	2	0.0221	0.0247	0.0141
2 qtrs	EMPCA	2	0.0033	0.0033	0.0034
	BPCA	2	0.0033	0.0034	0.0034
	SVDI	4	0.0053	0.0054	0.0042



Висновки:

У результаті кваліфікаційної роботи проведено порівняльний аналіз методів виділення факторів — EMPCA, TW, TP — для неповних змішаночастотних панельних даних. Найкращі результати продемонстрував метод EMPCA: він стабільно забезпечував найнижчі помилки прогнозування (RMSE, CRPS) як на коротких, так і довгих горизонтах, попри високу обчислювальну складність. Метод SVDI виявився ефективним для середніх горизонтів, а ВРСА — для довгострокових, але поступався за точністю в короткострокових прогнозах. На основі отриманих результатів EMPCA обрано основним методом для побудови моделей nowcasting у цій роботі.



Подальші напрями дослідження:

1. Інтеграція національних макропоказників
2. Використання нетрадиційних джерел
3. Застосування методів машинного навчання
4. Розробка адаптивних nowcasting-систем



Використана література:

- [1] Gary Koop, Stuart McIntyre, James Mitchell, Aubrey Poon, and Ping Wu. *Incorporating Short Data into Large Mixed-Frequency VARs for Regional Nowcasting* // Federal Reserve Bank of Cleveland Working Paper Series — May 2023
- [2] Jack Fosten, Ryan Greenaway-McGrevy. *Panel data nowcasting*. // *Econometric Reviews*, 41:7, 675-696 — 2022.
- [3] Mihnea Constantinescu, Kalle Kappne, Nikodem Szumilo. *The Warcast Index: Estimating Economic Activity without Official Data during the Ukraine War in 2022* // NBU Working Papers — March 2024