

ПОТОЧНЕ ПРОГНОЗУВАННЯ ВВП ЗА ДОПОМОГОЮ МОДЕЛЕЙ ФАКТОРІВ ЗІ ЗМІШАНОЮ ЧАСТОТОЮ ДЛЯ УКРАЇНСЬКИХ РЕГІОНІВ

Своєчасна оцінка регіональної економічної активності є ключовою для прийняття обґрунтованих рішень та реагування на кризові ситуації, особливо в економіках, де офіційна статистика публікується із суттєвими затримками. В Україні дані про валовий регіональний продукт (ВРП) оприлюднюються лише один раз на рік із запізненням до 16 місяців, що значно ускладнює моніторинг економічної ситуації в реальному часі. У цьому дослідженні запропоновано модель *Mixed-Frequency Factor-Augmented Vector Autoregression (MF-FAVAR)* для теперішнього прогнозування (*nowcasting*) квартального зростання ВРП для київського регіону шляхом поєднання річних, квартальних і місячних показників.

Запропонована структура інтегрує традиційні макроекономічні статистичні показники з високочастотними цифровими сигналами, отриманими з *Google Trends*, що дає змогу відстежувати зміни у споживчих настроях та поведінкових паттернах. Ці цифрові індикатори виступають проксі змін у споживчих настроях, намірах щодо витрат і очікуваннях на ринку праці, надаючи додаткову інформацію порівняно з офіційною статистикою, що публікується із затримкою. Зменшення розмірності даних здійснюється за допомогою сучасних методів факторної екстракції, розроблених для неповних і неузгоджених наборів даних, зокрема *Expectation-Maximisation Principal Component Analysis (EMPCA)*, *Bayesian PCA (BPCA)* та *Singular Value Decomposition Imputation (SVDI)*. Річний ряд ВРП було перетворено на квартальний за допомогою методу Дентона–Шолетта, що забезпечує узгодженість з офіційними підсумками.

Емпіричні результати показують, що факторна екстракція на основі EMPCA забезпечує найстабільніші та найточніші коротко-, середньо- та довгострокові прогнози. Зокрема, EMPCA досягає найменших значень середньоквадратичної похибки прогнозу (RMSFE) та безперервного рангового ймовірнісного показника (CRPS), що підтверджує її стійкість у умовах обмежених і зашумлених даних. Отримані результати свідчать, що моделі зі змішаною частотою та факторною структурою є ефективним інструментом для регіонального *nowcasting* за умов нестачі даних, що робить їх особливо релевантними для перехідних і кризових економік, таких, як Україна.

Ключові слова: MF-FAVAR, FAVAR, поточне прогнозування, EMPCA, GRP, RMSE, MAPE, Google Trends.

Вступ

Своєчасний моніторинг регіональної економічної активності набуває дедалі більшої ваги в сучасних макроекономічних дослідженнях та у сфері формування економічної політики. У періоди структурної нестабільності чи кризових явищ, зокрема під час війни, що триває в Україні, традиційні статистичні системи зазнають суттєвих затримок у публікації даних, що істотно ускладнює оцінювання економічної динаміки в режимі реального часу. Так, офіційна статистика валового регіонального продукту (ВРП) в Україні оприлюднюється з лагом до 16 місяців після звітного року, залишаючи органи державного управління, бізнес та наукову спільноту без актуальних індикаторів локальної еко-

номічної динаміки.

Задля подолання цього інформаційного розриву у сучасних прогнозних системах активно застосовують методи *nowcasting*. У концептуалізації Баньбура та ін. (2011, 2013)[1; 2], *nowcasting* означає використання неповних та асинхронних потоків даних для оцінювання поточного стану економіки. Інтегруючи високочастотні індикатори — щомісячні або щотижневі статистичні показники, результати опитувань чи цифрові поведінкові метрики, — такі моделі надають своєчасні сигнали економічної активності ще до оприлюднення офіційних релізів. Подібні підходи особливо цінні для економік, що стикаються з обмеженістю даних, частими ревізіями або перервами у статистичному спостереженні.

У контексті українських регіонів ці викли-

ки проявляються особливо гостро. Обмежена довжина часових рядів, асинхронність публікацій показників та структурні зміни, зумовлені війною, зумовлюють необхідність застосування стійких, гнучких та інформаційно-ефективних методів моделювання. Суміжні підходи прогнозування в умовах обмеженої інформації вже розглядалися у працях, що поєднують машинне навчання та економетричні гібридні моделі (Дрінь та ін. (2023)[19]). Нещодавня література також підкреслює ефективність змішаночастотних та фактор-розширених підходів. Зокрема, модель *Mixed-Frequency Vector Autoregression (MF-VAR)*, запропонована Куп та ін. (2023)[3], забезпечує системний підхід до роботи з незбалансованими наборами даних та інтеграції змінних різної частоти. Водночас MF-VAR стають нестабільними за великої кількості предикторів і коротких вибірок — що є типовим для регіональних досліджень.

Для подолання цих обмежень у цьому дослідженні застосовано модель *Mixed-Frequency Factor-Augmented Vector Autoregression (MF-FAVAR)*, яка поєднує переваги змішаночастотної інтеграції та факторного зниження розмірності. Такий підхід дозволяє стискати інформацію з широкого спектра економічних і цифрових індикаторів у невелику кількість латентних факторів, що відображають коваріації в регіональній економіці. Ці латентні фактори надалі формують динамічну систему VAR для отримання квартальних nowcast-оцінок ВРП на прикладі міста Київ.

Місто Київ обрано фокусним регіоном з трьох основних причин: (i) воно має найбільш повний набір регіональної статистики в Україні; (ii) його економічна структура є високодиверсифікованою, охоплюючи як промислові, так і сервісні сектори; (iii) цифрові індикатори, зокрема дані Google Trends, особливо інформативні для цього регіону, надаючи цінні проксі-змінні для вимірювання настроїв та поведінки споживачів.

Стаття робить внесок у розвиток досліджень субнаціонального nowcasting у трьох напрямках. По-перше, вона впроваджує сучасну модель MF-FAVAR в український контекст, демонструючи її працездатність в умовах екстремальної обмеженості даних. По-друге, порівнюються кілька просунутих алгоритмів факторної екстракції — Expectation-Maximisation PCA (EMPCA), Bayesian PCA (BPCA) та Singular Value Decomposition Imputation (SVDI) — з погляду стабільності та точності прогнозу. По-третє, оцінюється робастність запропонованого підходу за різних прогнозних горизонтів та конфігурацій даних.

Отримані результати свідчать, що фактор-

не зниження розмірності за допомогою ЕМРСА суттєво підвищує точність прогнозування, особливо на коротких і довгих горизонтах, водночас забезпечуючи стабільність в умовах нерегулярних чи неповних даних. Це підтверджує високий потенціал змішаночастотних фактор-розширених методів для моніторингу регіональної економічної динаміки у країнах з перехідною економікою та в умовах кризових викликів.

Подальша структура статті є такою: викладено теоретичні засади та представлено MF-FAVAR-модель; описано джерела даних та методи факторної екстракції; наведено емпіричні результати та оцінку прогнозної якості; подано ключові висновки та окреслено напрями майбутніх досліджень.

Попередні теоретичні засади

Nowcasting. Nowcasting означає оцінювання поточного стану економіки в режимі реального часу на основі неповних та асинхронних потоків даних. Такий підхід розроблено для подолання затримок офіційної статистики шляхом інтеграції високочастотних індикаторів (Баньбура та ін. (2013) [1]). Як підкреслюють Баньбура та ін. (2011) [2], nowcasting забезпечує короткострокові прогнози економічної активності ще до виходу офіційних макроекономічних показників, що є критично важливим у періоди швидких змін, кризових явищ або дефіциту даних.

У регіональному українському контексті — де доступність даних є обмеженою, а часові ряди порушені війною, — змішаночастотні підходи набувають особливої актуальності. Коор та ін. (2023) [3] пропонують великі змішаночастотні VAR-моделі (MF-VAR), які враховують короткі вибірки та явно розв'язують проблему «рваного краю» як на початку, так і наприкінці вибірки. Типовим викликом для України є те, що дані про ВРП публікуються лише щорічно і з істотним лагом (10–16 місяців). Для операціоналізації сигналів у режимі реального часу високочастотні індикатори (щомісячні/щотижневі) можна поєднувати з річними цільовими змінними за допомогою часової узгодженості та факторного стискання. У цьому дослідженні для Києва квартальні значення ВРП було отримано шляхом дезагрегації за методом Дентона–Шолетта ([7],[8]), а вимірність знижено завдяки факторній екстракції перед застосуванням VAR-моделювання.

Паралельна література формалізує обробку неузгоджених потоків даних у реальному часі. Динамічні факторні моделі (Dynamic factor

model – DFM), змішане вибірконе використання даних (Mixed Data Sampling – MIDAS) та MF-VAR автоматизують експертні оцінки за умов великих та асинхронних публікацій [1]. На багаторегіональному рівні панельний nowcasting підвищує стабільність та забезпечує крос-секційне використання інформації (Фостен та ін., (2022) [16]), що вказує на природне розширення підходу за межі одного регіону.

Класична VAR-модель. Векторна авторегресія (VAR) (Сімс (1980) [4]) описує K -вимірний вектор y_t у вигляді

$$y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + \varepsilon_t, \varepsilon_t \sim \mathcal{N}(0, \Sigma), \quad (1)$$

де $A_i \in \mathbb{R}^{K \times K}$, а ε_t є похибкою. VAR-моделі широко застосовують для прогнозування та структурного аналізу, однак вони потребують збалансованих панелей і містять $K^2 p$ параметрів, що призводить до надмірної параметризації у випадку невеликих регіональних вибірок.

Змішаночастотні VAR (MF-VAR). MF-VAR розширює класичні VAR для спільного використання змінних, що спостерігаються з різною частотою (щомісячно, щоквартально, щорічно), шляхом їх узгодження в регресійній або стан-просторовій постановці. Типові інструменти включають MIDAS-поліноми, фільтр Калмана або базове узгодження (Фороні та ін. (2013) [5]; Шорфхайде та ін. (2015) [6]). Попри ефективність у nowcasting, MF-VAR не має вбудованих механізмів зниження розмірності: за великої кількості високочастотних індикаторів виникає ризик надмірної параметризації та нестабільності, особливо за умов «рваних країв» (Куп та ін. (2023) [3]), характерних для українських регіонів.

Змішаночастотні FAVAR (MF-FAVAR)

Структура моделі. MF-FAVAR інтегрує механізм зниження розмірності у змішаночастотну VAR шляхом введення невеликого набору латентних факторів, що узагальнюють коваріації між багатьма спостережуваними індикаторами. Спостережувані змінні поділяють на високочастотні $x_t^{(H)} \in \mathbb{R}^{n_H}$ (наприклад, щомісячні макроекономічні та цифрові показники) та низькочастотні $y_t^{(L)} \in \mathbb{R}^{n_L}$ (наприклад, річний ВВП). Нехай $f_t \in \mathbb{R}^r$ — r латентних факторів, що еволюціонують як

$$f_t = \Phi_1 f_{t-1} + \dots + \Phi_p f_{t-p} + \nu_t, \nu_t \sim \mathcal{N}(0, Q), \quad (2)$$

де $\Phi_i \in \mathbb{R}^{r \times r}$. Змішаночастотні рівняння вимірювання пов'язують спостережувані змінні з

факторами:

$$x_t^{(H)} = \Lambda^{(H)} f_t + \eta_t^{(H)}, \quad (3)$$

$$y_t^{(L)} = \Lambda^{(L)} f_t + \eta_t^{(L)}, \quad (4)$$

де $\Lambda^{(H)}$ та $\Lambda^{(L)}$ — матриці навантажень, а $\eta_t^{(\cdot)}$ — унікальні збурення (idiosyncratic components). Квартальна цільова змінна ВРП (побудована нижче) входить до $y_t^{(L)}$ та прогнозується на основі f_t ; своєю чергою, фактори формуються в режимі реального часу завдяки більш оперативним $x_t^{(H)}$.

Позначимо $Z_t = [f_t^\top, y_t^{(L)\top}]^\top$. Тоді система MF-FAVAR може бути подана у вигляді:

$$Z_t = A_1 Z_{t-1} + \dots + A_p Z_{t-p} + \varepsilon_t, \varepsilon_t \sim \mathcal{N}(0, \Sigma). \quad (5)$$

Таке спільне представлення охоплює як латентні, так і спостережувані низькочастотні дані та об'єднується в єдину форму.

Переваги. MF-FAVAR забезпечує: (i) *зниження розмірності* (невелика кількість r замінює численні n_H фактори), (ii) *стійкість* до «рваних країв» та пропусків завдяки факторній екстракції та (iii) *узагальненість* через динаміку спільних компонент, що відфільтровує збурений шум [2].

Дані та методологія

Опис даних та часове узгодження.

Дослідження зосереджено на місті Київ. Щорічні дані ВРП (2004–2021) отримано з Державної служби статистики України [20] та дезагреговано у квартальний ряд за методом Дентона (1971) [7] і Шолетта (1984) [8]. Формально, процедура Дентона–Шолетта знаходить квартальний ряд $\{y_t\}_{t=1}^{4T}$, що мінімізує

$$\min_{\{y_t\}} \sum_{t=2}^{4T} (y_t - y_{t-1})^2$$

$$\text{s.t.} \quad \sum_{t=4(\tau-1)+1}^{4\tau} y_t = Y_\tau, \quad \tau = 1, \dots, T, \quad (6)$$

де Y_τ — це спостережувані річні підсумки. Оптимізація забезпечує плавність квартальних змін за збереження річної узгодженості. Це дозволяє побудувати гладкий квартальний ряд, узгоджений з річними орієнтирами, що створює основу для моделювання квартального зростання.

Високочастотні предиктори включають щомісячні макроекономічні ряди (наприклад, споживчі ціни, заробітна плата, обіг роздрібною торгівлі, будівництво, зайнятість) та індикатори Google Trends, що відображають динаміку за

ключовими словами, пов'язаними зі споживанням і настроями (зокрема, *інвестиції, інфляція, безробіття, заощадження*, пошукові запити щодо брендів). Щомісячні ряди конвертуються у квартальні шляхом підсумовування для потокових змінних та усереднення для запасів; річні змінні пролонгуються у межах року. Змінні логарифмуються за потреби та стандартизуються.

Залучення даних Google Trends мотивоване їхньою доведеною чутливістю до економічних шоків і поведінкових зрушень. Спираючись на Шмідт та Фозен (2009) [17] та Індако (2020) [18], були зібрані індекси інтенсивності пошуку для ключових слів, пов'язаних із споживанням («купити», «знижка», «подорож»), цінами («інфляція», «девальвація») та ринком праці («робота», «зарплата»). Ці індекси відображають миттєву суспільну увагу та очікування, які, як правило, передують офіційним опитуванням і цінній статистиці. Дані Google збиралися щомісячно, нормувалися до шкали 0 – 100 і агрегувалися у квартальні середні для узгодження з частотою даних ВРП.

Факторна екстракція за умов пропусків. З огляду на незбалансованість покриття та «рвані краї», розглянуто кілька методів на основі головних компонент:

ЕМРСА: expectation–maximisation PCA — ітераційно відновлює пропущені значення при одночасному оцінюванні навантажень і факторних балів;

ВРСА: баєсівська PCA з апіорними розподілами та автоматичним визначенням релевантності (Оба та ін. (2003) [9]);

SVDI: ітераційне низькорівневе відновлення матриці за допомогою усіченого SVD (Хансен (1987) [10]);

TW/TP: підходи Tall–Wide та Tall–Projection для «рваних країв» (Бай (2021) [11], Каган та ін. (2023) [12]).

У випадку панелі даних для Києва методи ЕМРСА, ВРСА та SVDI забезпечують стабільні фактори; методи TW і TP було відхилено через недостатню кількість повних «якірних» рядів і брак покриття у вікнах.

Оцінювання та схема перевірки. Латентні фактори ($r = 2$) виокремлюються з квартальної матриці предикторів X_t за допомогою кожного зі стійких методів (ЕМРСА, ВРСА, SVDI). VAR-модель для $[Y_t, f_t]$ із лагами $p \in \{2, 4\}$ утворює основу MF-FAVAR. Часове розбиття даних передбачає використання приблизно 80% спостережень для оцінювання та 20% для перевірки. Рекурсивні прогнози генеруються для горизонтів 2, 4 та 24 квартали. Точність прогнозів узагальнюється за трьома

стандартними метриками: RMSFE (точність у точковому вимірі), CRPS (калібрування розподілу) та SMAPE (обмежена відносна похибка) (Чен та ін. (2017) [13], Гнетінг та Рафтері (2006) [14], Хайндман та Келер (2007) [15]).

$$\text{RMSFE} = \sqrt{\frac{1}{H} \sum_{h=1}^H (\hat{y}_{T+h|T} - y_{T+h})^2}, \quad (7)$$

$$\text{CRPS}(\hat{F}, y) = \int_{-\infty}^{\infty} [\hat{F}(z) - \mathbb{I}\{z \geq y\}]^2 dz, \quad (8)$$

$$\text{SMAPE} = \frac{100\%}{n} \sum_{t=1}^n \frac{|\hat{y}_t - y_t|}{(|\hat{y}_t| + |y_t|)/2}. \quad (9)$$

де $\hat{F}$ — прогнозна функція розподілу ймовірностей.

Результати

Основні висновки. У різних прогнозних горизонтах MF-FAVAR на основі ЕМРСА забезпечує найстабільніші та найточніші прогнози. Для дуже коротких горизонтів (2 квартали) ЕМРСА досягає найнижчих значень RMSFE та CRPS; на середніх горизонтах (4 квартали) конкурентоспроможним є метод SVDI; на довгих горизонтах (24 квартали) ЕМРСА дещо перевершує ВРСА.

Як показано на рис. 1, на довгому горизонті у 24 квартали MF-FAVAR з факторною екстракцією ЕМРСА забезпечує найточніші прогнози як у точковому, так і у розподільчому вимірі. Траєкторія RMSFE та CRPS для довгого горизонту свідчить про невелику, але систематичну перевагу ЕМРСА над ВРСА, тоді як SVDI демонструє гірші результати, що узгоджується з підсумком у табл. 1.

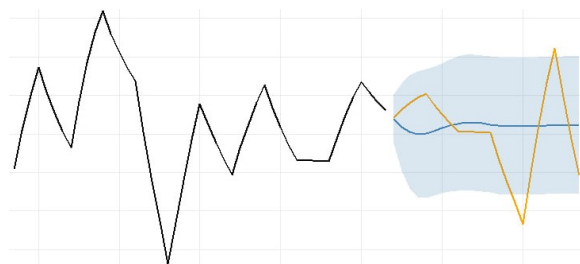


Рис. 1. Довгий горизонт (24 квартали)

Таблиця 1. Точність прогнозування за горизонтами (позавибіркове узагальнення)

Метод	Горизонт	RMSFE	CRPS
EMPCA	2Q	0.0033	0.0034
BPCA	2Q	0.0034	0.0034
SVDI	4Q	0.0167	0.0093
EMPCA	24Q	0.0208	0.0116

Примітка. Точність прогнозування зростає завдяки стабілізації динаміки факторів через зниження розмірності. EMPCA послідовно досягає найнижчих похибок на всіх горизонтах, особливо для 24-квартального горизонту, що ілюструється на рис. 1.

Простота моделі та стійкість. Перевірка на стійкість, яка виключає тижневі індикатори (розріджені та шумові) та зберігає базове ядро щомісячних і щорічних змінних, покращує точність, при цьому EMPCA знову демонструє перевагу. Це підкреслює важливість економічних наборів змінних у змішаночастотних моделях [2; 3].

Висновки та напрями подальших досліджень

Модель MF-FAVAR пропонує практичне рішення для регіонального nowcasting в умовах обмеженої доступності даних. Для Києва поєднання часово дезагрегованого ВРП з висо-

кочастотними макроекономічними та цифровими індикаторами, стислими за допомогою методів факторної екстракції на основі PCA, забезпечує точні квартальні nowcast-прогнози попри затримки офіційних публікацій. EMPCA виявляється послідовно стійким методом виділення факторів для панелей із «рваними краями», особливо на коротких горизонтах; SVDI демонструє конкурентні результати на середніх горизонтах; BPCA наближається до EMPCA на довших проміжках. Використання поведінкових даних із Google Trends підвищило чутливість короткострокових прогнозів, особливо у періоди структурних зламів та порушень статистичного збору, що підкреслює потенціал цифрових індикаторів для моніторингу регіональної економіки в реальному часі.

Подальші дослідження можуть: (i) поширити підхід на багаторегіональні панелі з метою використання крос-секційної інформації (Фостен та ін. (2022) [16]), (ii) інтегрувати додаткові високочастотні сигнали (мобільність, транзакції, соціальні медіа), а також (iii) здійснити порівняння з глибокими змішаночастотними моделями або гібридами MF-MIDAS. Отримані результати підтверджують, що факторне стискання є ключовим елементом для стабільного nowcasting за умов коротких та нерегулярних регіональних вибірок.

Список літератури

1. Bańbura M., Giannone D., Modugno M., Reichlin L. Nowcasting and the Real-Time Data Flow. ECB Working Paper No. 1564. 2013.
2. Bańbura M., Giannone D., Reichlin L. Nowcasting. The Oxford Handbook of Economic Forecasting. Oxford : OUP, 2011.
3. Koop G., McIntyre S., Mitchell J., Poon A., Wu P. Incorporating Short Data into Large Mixed-Frequency VARs for Regional Nowcasting. FRB Cleveland Working Paper No. 23-09. 2023.
4. Sims C. A. Macroeconomics and Reality. *Econometrica*. 1980. Vol. 48, No. 1. Pp. 1–48.
5. Foroni C., Marcellino M. A Survey of Econometric Methods for Mixed-Frequency Data. *Advances in Econometrics*. 2013. Vol. 31. Pp. 1–45.
6. Schorfheide F., Song D. Real-Time Forecasting With a Mixed-Frequency VAR. *Journal of Business & Economic Statistics*. 2015. Vol. 33, No. 3. Pp. 366–380.
7. Denton F. T. Adjustment of Monthly or Quarterly Series to Annual Totals. *Journal of the American Statistical Association*. 1971. Vol. 66, No. 333. Pp. 99–102.
8. Cholette P. A. Adjusting Sub-annual Series to Yearly Benchmarks. *Survey Methodology*. 1984. Vol. 10. Pp. 35–49.
9. Oba S., et al. A Bayesian Missing Value Estimation Method for Gene Expression Data. *Bioinformatics*. 2003. Vol. 19, No. 16. Pp. 2088–2096.
10. Hansen P. C. The Truncated SVD as a Method for Regularization. *BIT Numerical Mathematics*. 1987. Vol. 27, No. 4. Pp. 534–553.
11. Bai J., Ng S. Matrix Completion With Cross-Sectional and Serial Dependence. *Journal of Econometrics*. 2021. Vol. 222, No. 1. Pp. 413–430.
12. Cahan D., Foerster A., Sarte P.-D., Watson M. Nowcasting With Ragged-Edge Data: A ProjectionBased Approach. FRB Working Paper. 2023.
13. Chen C., Twycross J., Garibaldi J. M. A New Accuracy Measure Based on Bounded Relative Error. *PLoS ONE*. 2017. Vol. 12, No. 3. e0174202.
14. Gneiting T., Raftery A. E. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*. 2007. Vol. 102, No. 477. Pp. 359–378.
15. Hyndman R. J., Koehler A. B. Another Look at Measures of Forecast Accuracy. *International Journal of Forecasting*. 2006. Vol. 22, No. 4. Pp. 679–688.
16. Fosten J., Greenaway-McGrevy R. Panel Data Nowcasting. *Econometric Reviews*. 2022. Vol. 41, No. 7. Pp. 675–696.
17. Schmidt T., Vosen S. Forecasting Private Consumption: Survey-Based Indicators vs. Google Trends. Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI), Report No. 155. 2009.
18. Indaco A. From Twitter to GDP: Estimating Economic Activity from Social Media. Carnegie Mellon University in Qatar, Education City, Doha. 2020.
19. Дрін С., Крючкова А., Толочкова В. Прогностична модель для продукту без історії з використанням LightGBM: модель ціноутворення для нового продукту. *Могилянський математичний журнал*. 2023. Т. 6. С. 6–13.
20. Державна служба статистики України. Ста-

тистика валового регіонального продукту. URL: https://www.ukrstat.gov.ua/operativ/menu/menu_u/nac_r.htm

References

1. M. Bańbura, D. Giannone, M. Modugno and L. Reichlin, ECB Working Paper No. 1564 (2013).
2. M. Bańbura, D. Giannone and L. Reichlin, in: *The Oxford Handbook of Economic Forecasting* (OUP, Oxford, 2011).
3. G. Koop, S. McIntyre, J. Mitchell, A. Poon and P. Wu, FRB Cleveland Working Paper No. 23-09 (2023).
4. C. A. Sims, *Econometrica*. **48** (1), 1–48 (1980).
5. C. Foroni and M. Marcellino, *Advances in Econometrics*. **31**, 1–45 (2013).
6. F. Schorfheide and D. Song, *Journal of Business & Economic Statistics*. **33** (3), 366–380 (2015).
7. F. T. Denton, *Journal of the American Statistical Association*. **66** (333), 99–102 (1971).
8. P. A. Cholette, *Survey Methodology*. **10**, 35–49 (1984).
9. S. Oba, et al., *Bioinformatics*. **19** (16), 2088–2096 (2003).
10. P. C. Hansen, *BIT Numerical Mathematics*. **27** (4), 534–553 (1987).
11. J. Bai and S. Ng, *Journal of Econometrics*. **222** (1), 413–430 (2021).
12. D. Cahan, A. Foerster, P.-D. Sarte and M. Watson, FRB Working Paper (2023).
13. C. Chen, J. Twycross and J. M. Garibaldi, *PLoS ONE*. **12** (3), e0174202 (2017).
14. T. Gneiting and A. E. Raftery, *Journal of the American Statistical Association*. **102** (477), 359–378 (2007).
15. R. J. Hyndman and A. B. Koehler, *International Journal of Forecasting*. **22** (4), 679–688 (2006).
16. J. Fosten and R. Greenaway-McGrevy, *Econometric Reviews*. **41** (7), 675–696 (2022).
17. T. Schmidt and S. Vosen, *Forecasting Private Consumption: Survey-Based Indicators vs. Google Trends* (Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI), Report No. 155, 2009).
18. A. Indaco, *From Twitter to GDP: Estimating Economic Activity from Social Media* (Carnegie Mellon University in Qatar, Education City, Doha, 2020).
19. S. Drin, A. Kriuchkova and V. Toloknova, *Mohyla Mathematical Journal*. **6**, 6–13 (2023).
20. State Statistics Service of Ukraine, “Gross Regional Product Statistics”, available at: https://www.ukrstat.gov.ua/operativ/menu/menu_u/nac_r.htm (accessed 30 September 2025).

S. Drin, A. Zhuravlova, G. Kriukova

NOWCASTING REGIONAL GDP WITH MIXED-FREQUENCY FACTOR MODELS FOR UKRAINIAN REGIONS

Timely assessment of regional economic activity is essential for evidence-based policy and crisis response, particularly in economies where official statistics are released with significant delays. In Ukraine, regional Gross Regional Product (GRP) figures appear only once per year and with a lag of up to 16 months, severely limiting real-time monitoring. This study develops a *Mixed-Frequency Factor-Augmented Vector Autoregression (MF-FAVAR)* model to nowcast quarterly GRP growth for the Kyiv region by combining annual, quarterly, and monthly indicators.

The proposed framework integrates traditional macroeconomic statistics with high-frequency digital signals derived from Google Trends, capturing shifts in consumer sentiment and behavioural patterns. These digital indicators serve as real-time proxies for household sentiment, consumption intentions, and labour market expectations, offering a complementary perspective to lagged official statistics.

Empirical results demonstrate that EMPCA-based factor extraction delivers the most stable and accurate nowcasts across short, medium, and long horizons. In particular, EMPCA achieves the lowest Root Mean Squared Forecast Error (RMSFE) and Continuous Ranked Probability Score (CRPS), confirming its robustness under sparse and noisy conditions. The findings suggest that mixed-frequency factor-augmented frameworks offer a practical and efficient solution for regional nowcasting under data scarcity, making them especially relevant for transition and crisis economies such as Ukraine.

Keywords: MF-FAVAR, FAVAR, Nowcasting, EMPCA, GRP, RMSE, MAPE, Google Trends.

Матеріал надійшов 07.10.2025



Creative Commons Attribution 4.0 International License (CC BY 4.0)