

Поточне прогнозування ВРП макрорегіонів України

MF-FAVAR та SARIMAX на основі офіційних та альтернативних даних

Спеціальність: 113 «Прикладна математика»

Студент: Болотов Єгор (2 курс магістратури)

Керівник: Дрінь Світлана, канд. фіз.-мат. наук

Проблематика та Контекст Дослідження

-  **Інформаційний розрив війни:** З 2022 року збір офіційної статистики частково призупинено, а дані публікуються із критичними затримками.
-  **Проблема "рваного краю" (Ragged Edge):** Традиційні економетричні моделі втратили ефективність через нерівномірну наявність даних на момент прогнозування.
-  **Асиметричні шоки:** Війна спричинила різні наслідки: руйнування промисловості на Сході та Півдні, проти напливу ВПО та зростання споживання на Заході.
-  **Рішення (Nowcasting):** Оперативне прогнозування поточного стану регіональних економік у реальному часі з використанням альтернативних даних.

Мета та Завдання Дослідження



Головна Мета

Розробити, імплементувати та оцінити модель MF-FAVAR для точного прогнозування регіонального ВВП у реальному часі в умовах структурних шоків.



Збір Даних

Інтегрувати низькочастотні офіційні показники та високочастотні альтернативні (супутники NASA, Google Trends, нерухомість) у єдину матрицю (725 індикаторів, 12 макрорегіонів).



Моделювання та Оцінка

Вирішити проблему відсутності даних (EMPCA) та провести порівняльний аналіз багатоджерельної моделі із цифровим бенчмарком (SARIMAX).

Архітектура Використаних Моделей

MF-FAVAR (Багатоджерельна)

Векторна авторегресія змішаної частоти з факторним розширенням. Вирішує проблему розмірності.

EMPCA: Алгоритм для заповнення пропусків та вилучення латентних факторів.

Fernandez: Метод часової дезагрегації річного ВРП у місячний тренд.

Перевага: Фіксує реальні руйнування завдяки фізичним індикаторам.

SARIMAX (Цифровий Бенчмарк)

Авторегресійна модель, побудована виключно на аналізі цифрових слідів (Google Trends).

PCA: Використовує метод головних компонент для топ-15 пошукових запитів.

War Dummy: Містить фіктивну змінну структурного шоку 2022 року.

Перевага: Висока чутливість до змін настроїв у стабільних регіонах.

Дані: Офіційна Макростатистика

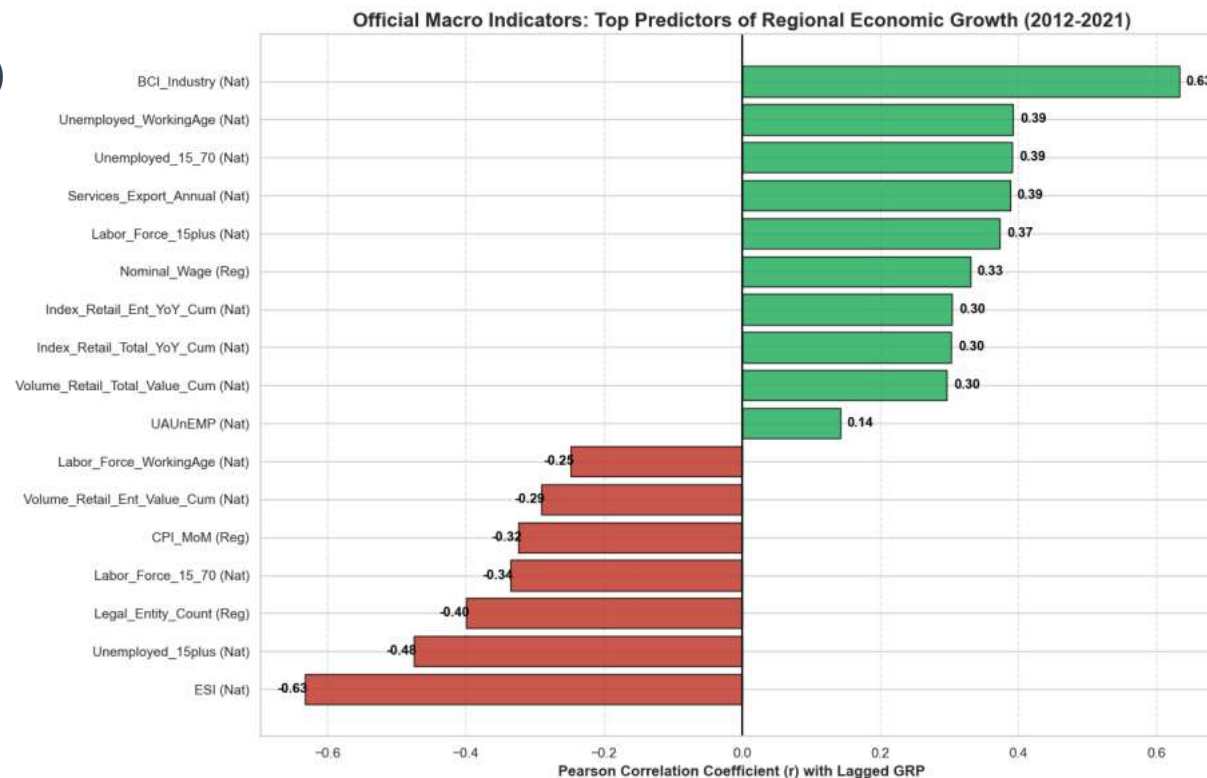
Основа моделі — набір з 91 національного та регіонального показника. Офіційні опитування (BCI, ESI) традиційно мають найвищу прогностичну силу.

Індекси: Споживчих цін (CPI), ділової впевненості (BCI, ESI).

Ринок праці: Номінальна зарплата, рівень безробіття.

Бізнес: Реєстрація ФОП, роздрібна торгівля.

Проблема: Під час війни ці дані публікуються із затримкою, що вимагає застосування альтернативних джерел.



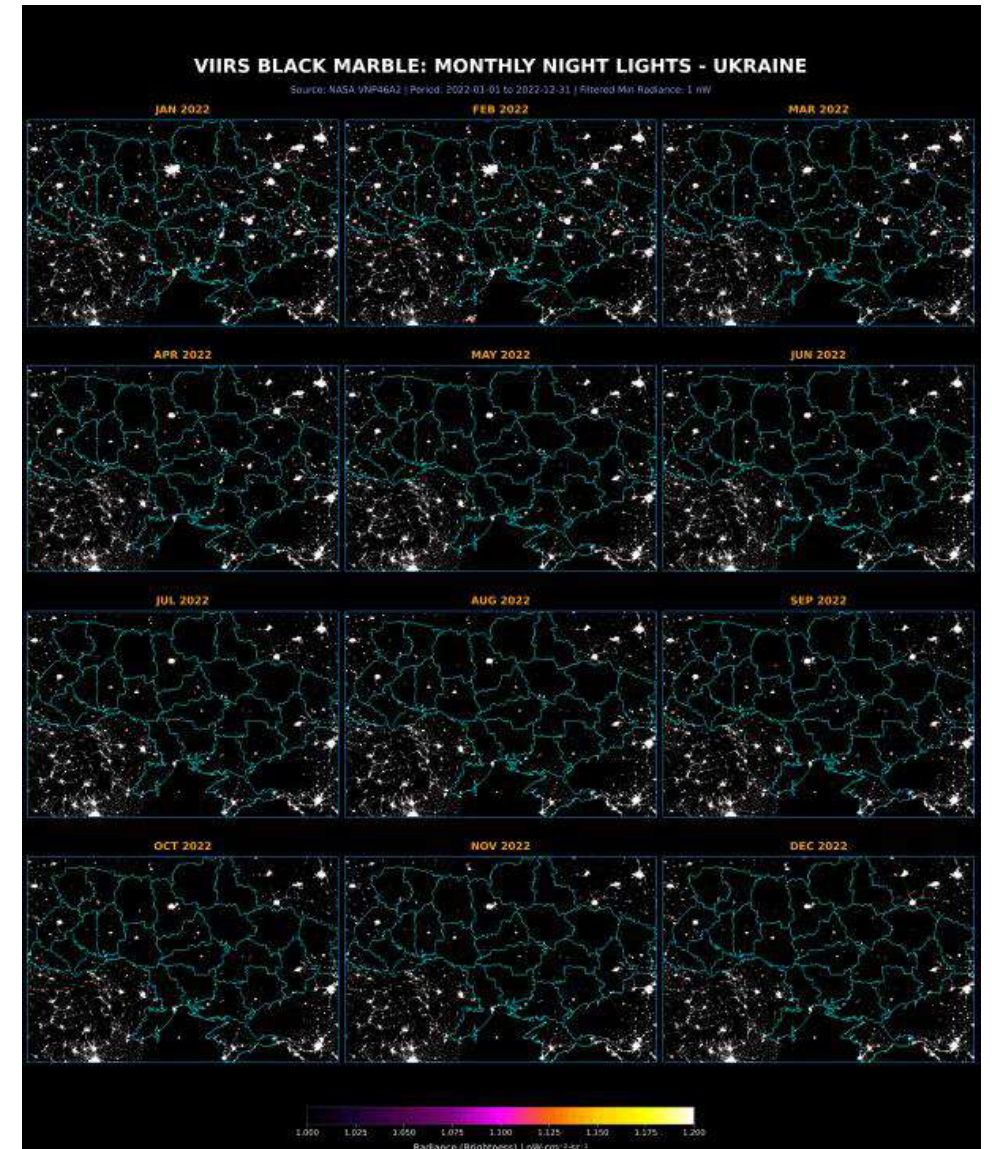
Дані: Фізична Телеметрія (NASA)

Супутникові дані NASA Black Marble (VIIRS) про інтенсивність нічного освітлення.

Об'єктивність: Незалежність від звітів на місцях; бачить реальну картину на окупованих територіях.

Індикатор: Сума світла сильно корелює з ВРП (до 0.48 у південних регіонах).

Фіксація шоків: Відображає колапс інфраструктури, блекауту та міграційні хвилі.

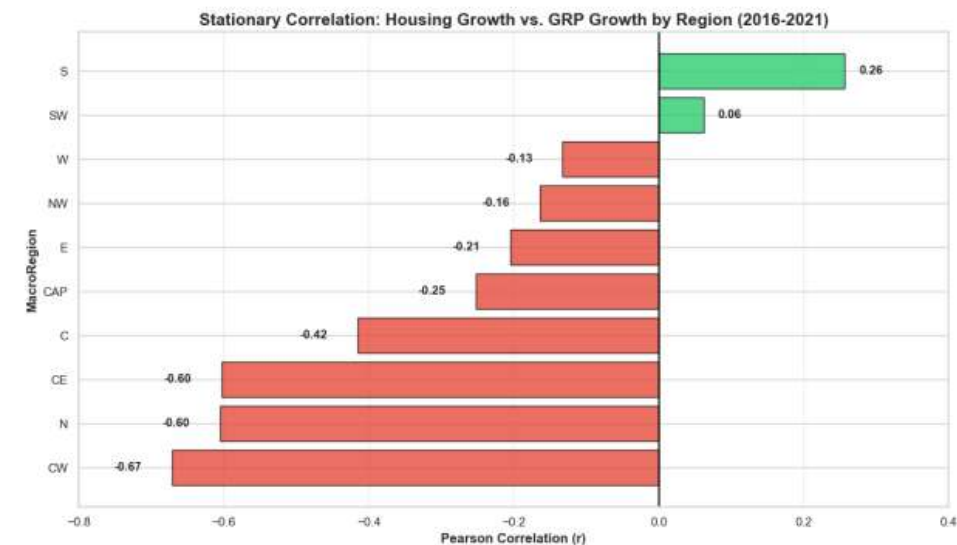
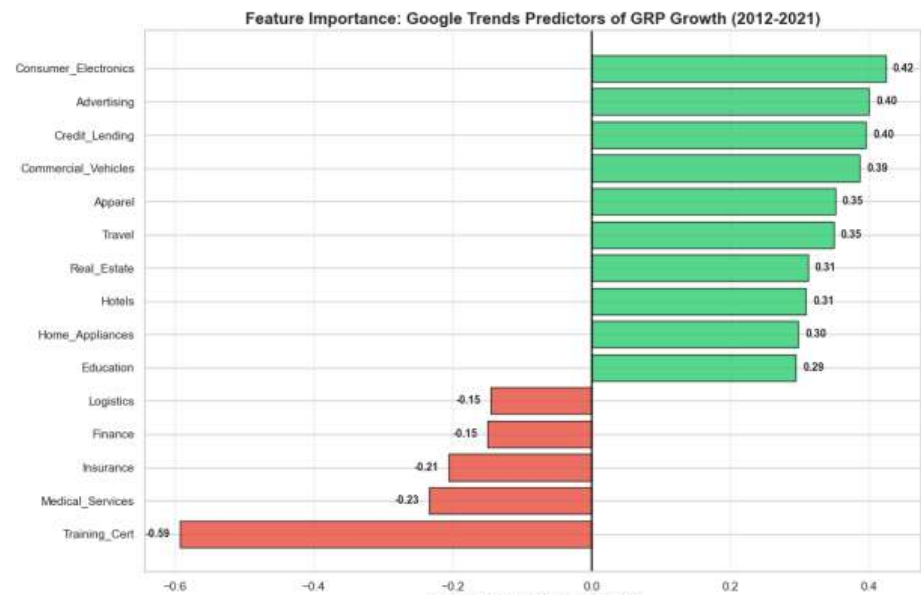


Дані: Цифрові Сліди та Нерухомість

Цифрові запити діють як миттєві індикатори попиту, обходячи інституційні "блекаути".

Google Trends (592 змінні): 52 категорії (рітейл, логістика, пошук роботи). Побутова електроніка та кредити — найсильніші позитивні предиктори.

M2Bomber (Нерухомість): Ціни в USD як індикатор, очищений від "номінальної ілюзії" інфляції. Фіксує збереження капіталу та попит ВПО.



Метрики Оцінки Точності Прогнозів

-  **RMSE (Root Mean Square Error):** Головна метрика. Квадратична похибка суворо карає за великі відхилення. Екстремальні промахи під час війни є критичними.
-  **MAE (Mean Absolute Error):** Середня абсолютна розбіжність без квадратичного зміщення від статистичних викидів.
-  **SMAPE (Symmetric Mean Absolute Percentage Error):** Масштабо-незалежна метрика для порівняння точності між регіонами різних розмірів.
-  **CRPS (Continuous Ranked Probability Score):** Оцінює точність імовірнісного розподілу (штрафує надмірну самовпевненість моделі).

Результати: MF-FAVAR vs SARIMAX (Out-of-Sample)

Макрорегіон	MF-FAVAR (RMSE)	SARIMAX (RMSE)	Інсайт (Різниця)
Донбас (DB)	0.1026	0.6618	Багатоджерельна модель (фізичні дані) стабільна. Бенчмарк колапсує.
Східний (E)	0.1018	0.1688	Перевага MF-FAVAR у зонах конфлікту та руйнувань.
Північно-Західний (NW)	0.0974	0.0371	Цифрова модель (SARIMAX) точніша у стабільних зонах.
Західний (W)	0.0602	0.0527	Висока чутливість Google Trends до споживання ВПО.

** Дані на основі бектестингу (2018-2021). MF-FAVAR перемагає в зонах конфлікту, спираючись на фізичну телеметрію.*

//

Цифрові пошукові метрики ідеально вловлюють наміри споживачів у стабільному середовищі, але повністю «сліпнуть» під час структурних шоків війни. Без заземлення на фізичну телеметрію (MF-FAVAR), бенчмарки хибно сприймають паніку за економічне

//

— *Продуктивна зворотність дослідження*

- ✓ Алгоритм EMPCA успішно стиснув 725 змінних у 2 латентні фактори без втрати даних.
- ✓ MF-FAVAR забезпечує надійні прогнози на коротких горизонтах (3-12 місяців).
- ✓ Створено готовий програмний пайплайн для моніторингу відбудови

Дякую за увагу!

Готовий відповісти на ваші запитання.