

Міністерство освіти і науки України
Національний університет «Києво-Могилянська академія»
Факультет інформатики
Кафедра математики

Кваліфікаційна робота
освітній ступінь – бакалавр

на тему: **«ПОРІВНЯННЯ КЛАСИЧНИХ VAR І MF-VAR
МОДЕЛЕЙ НА ПРИКЛАДІ ВВП ТА ІНФЛЯЦІЇ»**

Виконав: студент 4-го року
навчання
освітньої програми «Прикладна
математика»,
спеціальності 113 Прикладна
математика

Сірий Михайло Андрійович

Керівник: Дрінь С. С.

ст. викладач

Рецензент:

Кваліфікаційна робота захищена
з оцінкою _____

Секретар ЕК _____
(підпис)

«_____» _____ 20__р.

Міністерство освіти і науки України
Національний університет «Києво-Могилянська академія»
Факультет інформатики
Кафедра математики

ЗАТВЕРДЖУЮ

Зав.кафедри математики,
доцент, кандидат фіз.-мат. наук

_____ Чорней Р.К.
(підпис)

“ _____ ” _____ 2025

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ

для кваліфікаційної роботи
студенту 4-го курсу, факультету інформатики
Сірому Михайлу Андрійовичу

Тема: «Порівняння класичних VAR і MF-VAR моделей на прикладі ВВП»

Зміст кваліфікаційної роботи:

Анотація

1. Вступ

2. Огляд теоретичних основ класичних VAR моделей та MF-VAR моделей

3. Набір макроекономічних даних для аналізу та їх обробка

4. Побудова класичної VAR моделі

5. Побудова MF-VAR моделі

6. Порівняти якість прогнозів ВВП, отриманих за допомогою VAR та MF-VAR моделей

Висновки

Список літератури

Дата видачі “ _____ ” _____ 2025 Керівник _____
(підпис)

Завдання отримав _____
(підпис)

Графік підготовки кваліфікаційної роботи до захисту

Графік узгоджено «_____» _____ 2025р.

Науковий керівник _____
(ПІБ)

Виконавець кваліфікаційної роботи _____
(ПІБ)

№ з/п	Перелік робіт	Термін виконан- ня етапу	Підпис наукового керівника	Дата озна- йомлення наукового керівника	Примітки
1.	Вибір теми, затвердження її на засіданні кафедри та закріплення наукового керівника.	Жовтень 2025	Жовтень 2025		
2.	Вивчення джерел літератури, збір та узагальнення даних	Жовтень 2025 – Листопад 2025	Листопад 2025		
3.	Складання плану кваліфікаційної роботи та узгодження з науковим керівником.	Листопад 2025	13.12.2025		
4.	Написання розділів кваліфікаційної роботи.	Грудень 2025 – Лютий 2026			
5.	Завантаження чорнової версії кваліфікаційної роботи.	15.03.2026			
6.	Відгук від наукового керівника про кваліфікаційну роботу.	20.03.2026			
7.	Попередній захист кваліфікаційної роботи.	02.04.2026 – 04.04.2026			
8.	Фінальна версія кваліфікаційної роботи	20.05.2026			
9.	Захист кваліфікаційної роботи.	28.05.2026 – 29.05.2026			

Зміст

Анотація	5
Примітки	5
1 Вступ	6
1.1 Актуальність	6
1.2 Мета і завдання дослідження	6
2 VAR модель	8
2.1 Математична форма VAR моделі	8
2.2 Принципи роботи VAR-моделі	8
2.3 Приклад авторегресійної моделі	9
2.4 Застосування в макроекономіці	10
3 MF-VAR модель[1]	11
3.1 Байєсова ймовірність[11]	11
3.2 Групування змінних	13
4 Робота з даними	14
4.1 Опис даних	15
5 Результати VAR моделі	17
5.1 Результати для ВВП України	18
5.2 Результати для інфляції України	19
5.3 Імпульсно-перехідні функції (IRF)	21
6 Результати BVAR моделі	23
6.1 Коефіцієнти для BVAR моделі	23
6.2 Прогнози BVAR моделі	26
6.3 Імпульсно-перехідні функції	27
7 Результати моделі для змішано-частотних даних	29
7.1 Механіка роботи	29
7.2 Коефіцієнти моделі	30

7.3	Прогнози на майбутні періоди та поточне прогнозування	32
8	Результати моделей	35
9	Висновки	36
10	Додатки	38

Анотація

Регіональні економічні дані стають все дедалі цікавішими для аналізу та поточного прогнозування (nowcast). Але вони мають короткі часові ряди, що ускладнює побудову векторно-авторегресійної моделі, адже вони матимуть малу точність. Також на це впливає затримка публікації даних. В кращому випадку регіональні економічні показники публікуються кожен квартал, в гіршому випадку кожен рік. Це приводить до того, що в нашій моделі є багато пропущених даних і до погіршення результатів моделі.

У цій роботі ми працюватимемо з регіональними даними по українським областям. Ці дані мають пропуски по деяким кварталам через затримку публікації даних. Саме тому в цій роботі ми будемо працювати з моделями змішаних частотностей (MF-VAR). Ми вирішуємо цю проблему “нерівних країв” різними методами і в кінці дослідження порівняємо їх всіх між собою за метриками MSE та $RMSE$.

Примітки

Для оформлення роботи використовувались методи штучного інтелекту.

1 Вступ

1.1 Актуальність

У зв'язку з військовим станом в якому перебуває Україна, ми дуже залежні від макроекономічних показників таких як ВВП чи інфляція, які впливають на ріст чи падіння економіки. А це у свою чергу впливає на рівень доходу громадян, рівень підтримки Збройних Сил України та економічну стабільність держави в цілому. Саме тому питання передбачення макроекономічних показників є дуже важливою задачею для українських фінансистів сьогодні. І тому найголовнішою задачею є побудова якнайточнішою моделі, яка здатна буде передбачати на регіональному рівні ВВП та інфляцію.

1.2 Мета і завдання дослідження

Метою кваліфікаційної роботи є порівняльний аналіз класичних векторних авторегресійних моделей (VAR) та моделей зі змішаними частотностями (MF-VAR) у задачах прогнозування ВВП, а також оцінка переваг використання різночастотної інформації для підвищення точності коротко- та середньострокових прогнозів.

Завдання дослідження

1. Провести огляд теоретичних основ класичних VAR моделей та їх застосування у макроекономічному прогнозуванні ВВП.
2. Розглянути методологію MF-VAR моделей, зокрема принципи інтеграції змінних з різною частотою спостережень та їх економетричні припущення.
3. Сформувати набір макроекономічних даних для аналізу, що включає квартальні показники ВВП та місячні індикатори економічної активності (промислове виробництво, інфляція, зайнятість, роздрібна торгівля тощо).
4. Провести попередню обробку даних, включаючи трансформації, сезонне коригування та перевірку стаціонарності часових рядів.

5. Побудувати класичну VAR модель на основі агрегованих даних з однаковою частотою.
6. Побудувати MF-VAR модель з використанням квартальних і місячних змінних без попередньої агрегації високочастотних даних.
7. Порівняти якість прогнозів ВВП, отриманих за допомогою VAR та MF-VAR моделей, з використанням стандартних критеріїв точності прогнозування (RMSE, MAE тощо).
8. Проаналізувати чутливість результатів до вибору лагової структури та набору пояснювальних змінних.
9. Сформулювати висновки щодо доцільності застосування MF-VAR моделей для макроекономічного аналізу та прогнозування ВВП.

Очікувані результати

У результаті виконання кваліфікаційної роботи очікується отримання порівняльної оцінки ефективності класичних VAR та MF-VAR моделей у прогнозуванні ВВП. Робота має продемонструвати, чи забезпечує використання різночастотних даних покращення прогнозної здатності моделей, а також визначити умови, за яких MF-VAR підхід є методологічно та практично доцільним для макроекономічних досліджень.

2 VAR модель

Векторна авторегресійна модель (VAR) [5] — це статистичний інструмент, який використовується для дослідження динамічних зв'язків між кількома змінними часових рядів. У VAR-моделях береться взаємозв'язок одночасно від декількох змінних. Досягається це тим, що кожна змінна представлена як функція, яка залежить не лише від її попередніх значень, а й минулих значень інших змінних. У VAR-моделі кожна змінна регресує за власними відставаними значеннями та значеннями інших змінних у моделі.

2.1 Математична форма VAR моделі

VAR-модель математично представлена як система рівнянь, де кожне рівняння описує поведінку однієї змінної як функцію власних відставаних, на якийсь порядок p , значень. Це означає, що p в моделі VAR(p) позначає від якої кількості кроків залежить наша модель. Математично це описується так:

$$Y_t = c + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p} + \epsilon_t$$

Де:

- Y_t : значення часового ряду в момент часу t .
- c : вектор констант.
- Φ_p : авторегресійні коефіцієнти для лагу p .
- Y_{t-p} : значення часових рядів з лагами $1, 2, \dots, p$ до моменту часу t .
- ϵ_t : шоки або по-іншому випадкові стрибки.

2.2 Принципи роботи VAR-моделі

1. **Лінійність:** залежності між змінними мають бути лінійні.
2. **Стаціонарність:** часові ряди мусять бути стаціонарними, тобто детрендованими.

3. **Немає ідеальної мультиколінеарності:** тобто змінні не залежать одна від одної повністю лінійно.
4. **Немає автокореляції:** залишки не послідовно корелюють.
5. **Гомоскедастичність:** дисперсія залишків постійна.
6. **Достатня спостережуваність:** дані для оцінки параметрів адекватні і в моделі немає викидів.
7. **Слабка екзогенність:** тобто фактори, які задаються поза моделлю, але таких факторів мало.

2.3 Приклад авторегресійної моделі

Припустимо, ми маємо модель з двох незалежних змінних: ВВП та інфляції. Система рівнянь для VAR-моделі з кроком 1 (тобто кожна змінна залежить лише від одного попереднього кроку) виглядатиме, як:

$$\begin{cases} GDP_t^1 = c_1 + a_{11}GDP_{t-1} + a_{12}INF_{t-1} + \epsilon_{1,t} \\ INF_t^2 = c_2 + a_{21}GDP_{t-1} + a_{22}INF_{t-1} + \epsilon_{2,t} \end{cases}$$

$$\Phi_p = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

Тобто модель ВВП залежить від значень ВВП та інфляції на попередньому кроці та “білому” шумі. Для моделей, які залежать від більшої кількості кроків (лагів) додається добуток матриці авторегресійних коефіцієнтів Φ_p . Де коефіцієнти Φ_p це:

- a_{11} : показує як показник ВВП на попередньому кроці впливає на теперішній ВВП.
- a_{12} : показує як показник ВВП на попередньому кроці впливає на теперішню інфляцію.
- a_{21} : показує як попередня зміна інфляції впливає на теперішній ВВП.

²GDP тут позначає ВВП

²INF тут позначає інфляцію

- a_{22} : показує як попередня зміна ВВП впливає на теперішню інфляцію.

2.4 Застосування в макроекономіці

VAR-моделі є надзвичайно-важливими в макроекономіці, адже окрім звичайних регресійних моделей, які здатні передбачати значення на крок вперед, вони дозволяють робити короткостроковий прогноз (nowcast), що є дуже впливовим аспектом, адже дозволяє завдяки цьому корегувати наші прогнози макроекономічних показників на поточний період часу.

Також одним з найважливіших застосувань виділяють **аналіз імпульсних відгуків (IRF)**. Він дозволяє побачити як аномальне значення в одній змінній вплине на інші змінні протягом наступних періодів. Наприклад, за допомогою можна відстежити наскільки впаде ВВП протягом найближчих майбутніх періодів при теперішньому збільшенні тарифів.

Щоб знайти графіки імпульсних відгуків перепишемо VAR модель у форму VMA[8] (Vector moving average):

$$Y_t = \mu + \sum_{i=0}^{\infty} \Phi_i \epsilon_{t-i}$$

В цьому рівнянні Φ_i представляє собою матрицю імпульсних відгуків, які ми можемо бачити на графіках нижче.

3 MF-VAR модель[1]

VAR-моделі тренуються лише на повних датасетах, тобто тих, де немає пропущених значень. Але зазвичай, регіональні економічні показники публікуються рідше за національні, наприклад, там де є щоквартальні національні показники, в регіональних даних є лише річні. Таким чином в датасеті є пропущені значення по регіональним показникам. Вирішенням цієї проблеми є VAR моделі змішаних частотностей. Також ця модель дозволяє працювати з даними, які мають “рваний” край, тобто деякі показники закінчуються раніше за інші. Вирішенням цієї проблеми є фільтр Калмана, який створює приховані стани, використовуючи кореляцію з існуючих показників. Цей вид прогнозування називається **nowcasting**, тобто прогнозування на поточний період, заповнюючи пропуски в даних. У свою чергу, MF-VAR моделі є ефективнішими за звичайні VAR моделі, адже у VAR-моделях застосовують просту інтерполяцію, яка не завжди є вирішенням проблем з генерацією даних, на відміну від MF-VAR моделей.

3.1 Байєсова ймовірність[11]

Теорема Байєса є одним з найкращих та найвдаліших прикладів застосування VAR моделей зі змішаними частотностями, адже теорема описує умовну ймовірність події, а в нашому випадку це ймовірність події, значення якої ми не знаємо. В загальному виді формула така:

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)}$$

- **H**: гіпотеза про дані. Тобто наші дані, яких немає в моделі, але ми хочемо їх нагенерувати.
- **E**: реальні дані, які присутні в нашому датасеті.
- **P(H)**: апіорна ймовірність події H. Перефразовуючи на проблематику моделі, це наші припущення про дані H.
- **P(E|H)**: ймовірність події H за умови E, також називають як функцію

правдоподібності. Це визначає наскільки підібрана модель відповідає реальним даним.

- $P(H|E)$: апостеріорна ймовірність або результат моделі. Вона тим, що ми хочемо знати ймовірність гіпотези, яку ми поставили за умових реальних даних.

Minnesota prior

В моделях, де ми маємо мало спостережень, тренуючи MF-VAR модель, ми скоріш за все стикнемось з перенавчанням моделі. Це з-за того, що даних відносно мало і модель починає шукати закономірності, де їх немає. Вирішенням цієї проблеми є Міннесотська апріорна ймовірність. Воно не намагається шукати закономірності, а йде шляхом випадкового блукання. Основні припущення:

1. Значення змінної в момент часу t це значення цієї змінної в момент часу $t - 1$ плюс “білий” шум.
2. Вплив далеких лагів є несуттєвим.
3. Вплив інших змінних на спостережувану слабший, ніж вплив самої на себе.

Самі пріори задаються наступним чином[2]:

$$V_{i,kk} = \begin{cases} \frac{k_1}{7^2}, \text{ де } k_1 \text{ це стиснення для власних лагів змінної} \\ \frac{k_2 s_i^2}{l^2 s_j^2}, \text{ де } k_2 \text{ це стиснення для інших лагів змінної} \\ \frac{k_3 s_i^2}{s_j^2}, \text{ де } k_3 \text{ це стиснення для лагів екзогенних змінних} \\ k_4 s_i^2, \text{ де } k_4 \text{ це стиснення для константи} \end{cases}$$

Так ми маємо формули для розрахунку усіх параметрів для моделі, які дозволяють точніше налаштовувати модель. Завдяки цьому методу ми робимо модель простішою, яка не перенавчається і не залежить від дуже великої кількості параметрів.

Алгоритм Гіббса

Наш датасет містить багато національних та регіональних даних, роблячи його дуже важким для підбору параметрів для моделі. В цій задачі нам допомагає алгоритм Гіббса. Він працює ітеративним методом в такі кроки:

1. Оцінка даних на припущенні, що ми знаємо модель.
2. Оцінюємо параметри моделі.
3. Повторимо процес декілька тисяч разів.

Таким чином ми отримуємо апостеріорний розподіл за допомогою якого можна знайти найбільш ймовірне значення для пропущених даних.

3.2 Групування змінних

Цей метод працює, коли в нас деякі змінні є місячними, а деякі є квартальними. Ми представляємо Y_t як вектор-стовпчик з 4 змінних, де одна квартальна змінна, а інші 3 є місячними. Тобто,

$$Y_t = \begin{bmatrix} x_{1t} \\ x_{2t,m1} \\ x_{2t,m2} \\ x_{2t,m3} \end{bmatrix}$$

І далі ми записуємо звичайне рівняння векторної авторегресії:

$$Y_t = c + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \dots + \Phi_p Y_{t-p} + \epsilon_t$$

Таким чином ми уникнули складних перетворень моделі, щоб нагенерувати значення та не застосовували складний метод Калмана. Але недоліки полягають в тому, що метод групування змінних добре працює для змінних, які мають великий часовий ряд. А в нашому прикладі з регіональними даними це створює проблеми, адже більшість даних там з 2015 до 2025.

4 Робота з даними

Для даних ми візьмемо макроекономічні показники з Державної служби статистики[12]. Основні дані з якими ми будемо звіряти зі статєю[4]. Основна проблема полягає в тому, що макроекономічні дані публікуються з великою затримкою, а через специфікації VAR моделі, ми не можемо працювати з даними, які мають пропуски. Тому модель буде складатись з кількох лише показників, бо додавання до моделі ще більшої кількості показників, задля збільшення якості прогнозування, призведе до перенавчання моделі і нездатності адекватно робити прогнозування. Але ще перед тренуванням моделі необхідно зробити наступні трансформації для даних:

1. **Логарифмування.** Для даних, які мають числові значення, а не відсоток, необхідно робити логарифмування, щоб зменшити абсолютні значення. Це дуже важливо, адже якщо цього не зробити, то похибка моделі може становити декількох відсотків, що відчутно для таких макроекономічних показників як ВВП.
2. **Стаціонарність.** Це є однією з найважливіших трансформацій для даних моделі. Адже стаціонарність ряду означає постійні середнє та дисперсію, що означає, що модель не буде коливатись з часом. Зазвичай для усунення стаціонарності беруть перші різниці між поточним та минулим значенням. Далі за тестом Дікі-Фулера перевіряють цей ряд на нормальність і якщо треба ще раз беруть перші різниці.
3. **Сезонність.** Це коливання, які відбуваються через певні проміжки часу. Для часових рядів це є дуже великою проблемою, адже модель може зазвичай цей паттерн і постійно слідувати лише цьому тренду. Таким чином ми маємо недонавчання моделі з яким борються шляхом синусоїдальних рядів або моделі SARIMA. Також є метод фіктивних змінних, він працює за методом додавання нових бінарних змінних до моделі, задля усунення тренду. Метод фіктивних змінних працює лише з категоріальними змінними, але для часових рядів ми можемо представити 4 квартали як 4 різні категорії. Це дозволяє нам легше зловити взаємозв'язок між часом і даними для виокремлення тренду.

Після цих трансформацій в нас будуть чисті дані на яких можна тренувати VAR модель. Отримані результати будуть показувати не абсолютне значення, а будуть залишками. Тому після отримання прогнозів необхідно їх повернути до звичайного розміру через додавання до останнього значення суми залишків та піднесення його до експоненти, якщо було застосовано логарифмування.

4.1 Опис даних

- **ВВП України.** Позначає валовий продукт України в мільйонах гривнях. Присутній щоквартально починаючи з 2015 року.
- **Індекс цін.** Позначає зміну цін порівняно з попереднім місяцем. Присутній щомісячно з серпня 1991 року.
- **Індекс зарплат.** Зміна середньої зарплати окремо по Україні та Київській області.
- **Експорт і імпорт товарів.** Експорт товарів це процес вивезення товарів в інші країни. Оцінюється в мільйонах гривнях. Імпорт товарів це закупівля товарів з інших країн. Оцінюється в мільйонах гривнях. Обидва показники присутні в датасеті моделі, починаючи з першого кварталу 2015 року.
- **Зайнятість.** Вимірює кількість працездатних людей, які мають роботу. Дані присутні по Україні та Київській області, публікуються з першого кварталу 2015 року.
- **Індикатори ділової впевненості.** Це композитні показники, які базуються на опитуваннях підприємств в різних галузях. Дані присутні по фінансовій сфері, сфері послуг та будівництві. Публікуються з першого кварталу 2015 року.
- **Житлова площа.** Загальна площа двокімнатних квартир по Україні загалом та Київській області. Дані публікуються з першого кварталу 2018 року.

- **Регіональний валовий продукт.** Позначає валовий продукт Київської області. Публікується з 2004 року щорічно.

5 Результати VAR моделі

Перша модель на якій були натреновані дані це проста VAR-модель, натренована з-за допомогою пакету *statsmodels.tsa.api*[10]. VAR-модель не вміє працювати з пропущеними даними, тому основною проблемою цієї моделі стало те, що занадто мало спостережень було по кожному параметру. Також велика проблема полягала в тому, що регіональний валовий продукт публікується лише раз на рік. Вирішенням цієї проблеми став метод інтерполяції, що з'єднує плавно точки. Цей метод хороший, якщо нам просто треба значення, але у великих моделях таке не використовують, бо по кварталам можуть бути неочікувані підйоми або падіння значень, але інтерполяція цього виявити не зможе.

Натренувавши модель, маємо таку кореляційну матрицю для залишків: На цій матриці видно, що нема дуже сильних зв'язків між якимись двома

	gdp_ua	enterprises	kyiv_gdp	sal_ind_kyiv	infl_ukr
gdp_ua	1.000000	0.267030	-0.589607	-0.529271	-0.082057
enterprises	0.267030	1.000000	-0.668464	-0.571441	0.692032
kyiv_gdp	-0.589607	-0.668464	1.000000	0.853162	-0.302650
sal_ind_kyiv	-0.529271	-0.571441	0.853162	1.000000	-0.580071
infl_ukr	-0.082057	0.692032	-0.302650	-0.580071	1.000000

Табл. 1: Кореляційна матриця для залишків

змінними. ВВП України найбільше корелює з валовим продуктом Київської області, що логічно. Але кореляція від'ємна, що можна пояснити відтоком капіталу до області. Також причина чому кореляція між Київською областю і Україною від'ємна в тому, що Київщина, де розташовано більшість ІТ-підприємств, швидше реагує на глобальні зміни. Та сама причина негативної кореляції і для зарплат Київської області. Зміни кількості підприємств негативно корелюють з усіма Київськими показниками, що можна пояснити тим, що дуже не рентабельно відкривати бізнес через високі ціни. В той же час збільшення підприємств збільшує зміну інфляції.

5.1 Результати для ВВП України

Прогноз VAR-моделі для ВВП України: $\text{gdp_ua} = -0.24 + 0.37Q2 + 0.45Q3 + 0.29Q4 - 0.06 \cdot \text{gdp}_{t-1} + 0.53 \cdot \text{ent}_{t-1} + 0.28 \cdot \text{kyiv_g}_{t-1} + 0.003 \cdot \text{sal}_{t-1} + 0.005 \cdot \text{infl}_{t-1} - 1.57 \cdot \text{empl}_{t-1}$.

Таким чином видно, ми побудували VAR модель для ВВП. Перед тренуванням моделі, дані були розділені так, щоб останній рік з 4 кварталів йшов на тестування з передбаченими даними моделі. На наступних малюнках результати роботи моделі для ВВП України.

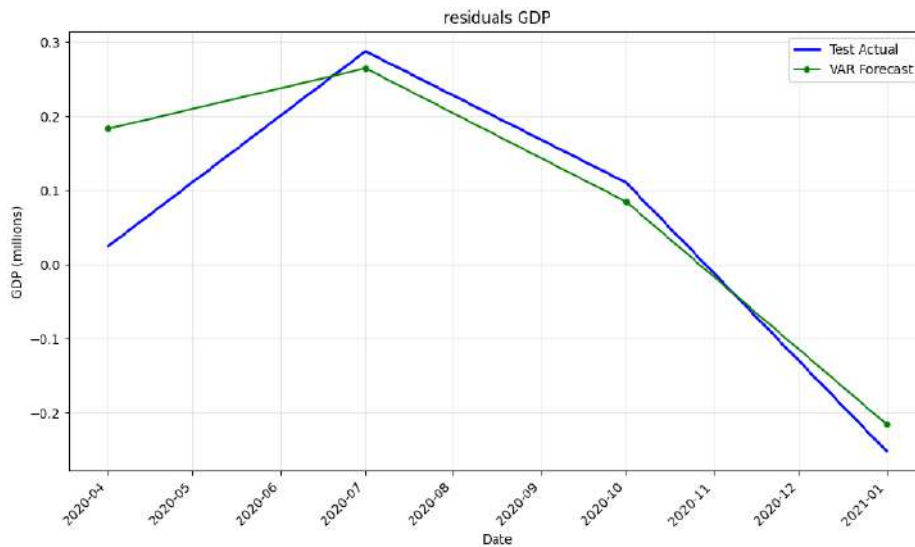


Рис. 1: Графік залишків для ВВП

На малюнку можна бачити графік залишків для ВВП України. Для інтерпретації це зміна ВВП між двома кварталами. Можна бачити, що модель ідеально впіймала тренд та сезонність. Єдина мала похибка полягає в першій точці, де модель спрогнозувала більшу набагато зміну значень для другого кварталу. Модель навчалась на даних, де не було викидів через політичні події, а перша прогнозована точка позначає другий квартал 2020 року. Саме тоді була епідемія ковіду і були заборони на масові скупчення людей, саме тому люди залишались вдома і витрачали менше тим самим держава не отримувала прибуток.

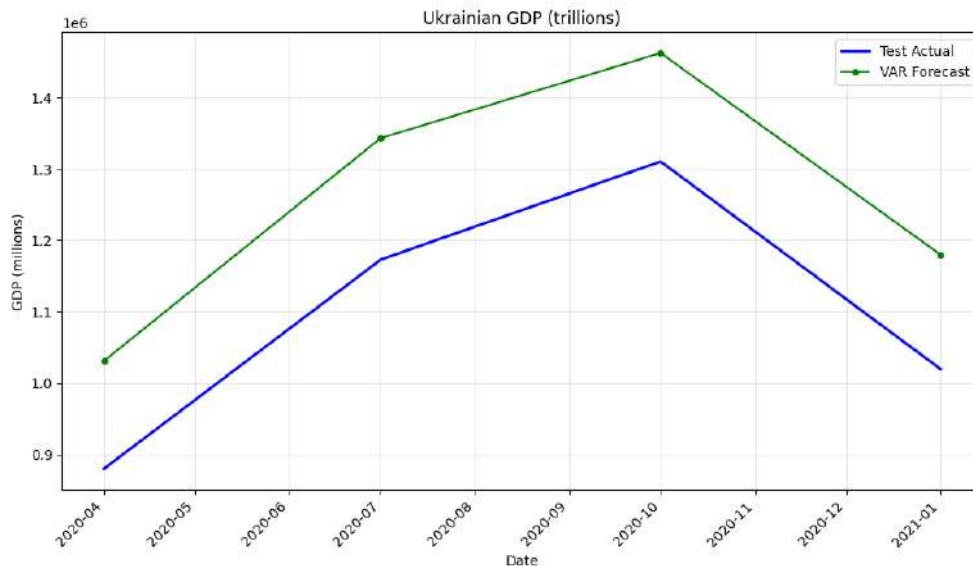


Рис. 2: ВВП України

На графіку видно, що модель дуже добре виймала тренд та сезонність, що було вже сказано в попередньому пункті. Але тут ми вже повернулись до нормальних значень методом додавання до останнього значення суми залишків та піднесенням до експоненти. Таким чином ми повернули масштаб оригінальних значень. Модель трохи переоцінила ВВП, але всеодно результат є дуже хорошим зважаючи на такий малий датасет та простоту моделі.

5.2 Результати для інфляції України

Прогноз для VAR моделі для інфляції: $\text{infl_ua} = 2.14 - 11.09 \cdot Q2 + 2.33 \cdot Q3 + 6.57 \cdot Q4 - 38.15 \cdot \text{gdprua}_{t-1} + 43.86 \cdot \text{ent}_{t-1} - 9.04 \cdot \text{gdpkyiv}_{t-1} + 0.19 \cdot \text{sal}_{t-1} - 0.22 \cdot \text{gdprua}_{t-1}$. Серед усіх змінних модель виділила декілька, які є статистично-значущими ($pvalue < 0.05$). Це Q2, Q4, gdprua та enterprises. Це логічно, адже показники за квартали явно показують коли найімовірніше зросте інфляція, таке саме можна сказати про ВВП. Щодо показника кількості підприємств, то модель неочікувано сказала, що цей показник є значущим та має найбільший коефіцієнт з усіх інших. Це означає, що при збільшенні кількості підприємств збільшується і інфляція. Напрямку знайти причину цього складно, адже не володію обширними знаннями з макрофінансів, але це цікава ситуація для аналізу.

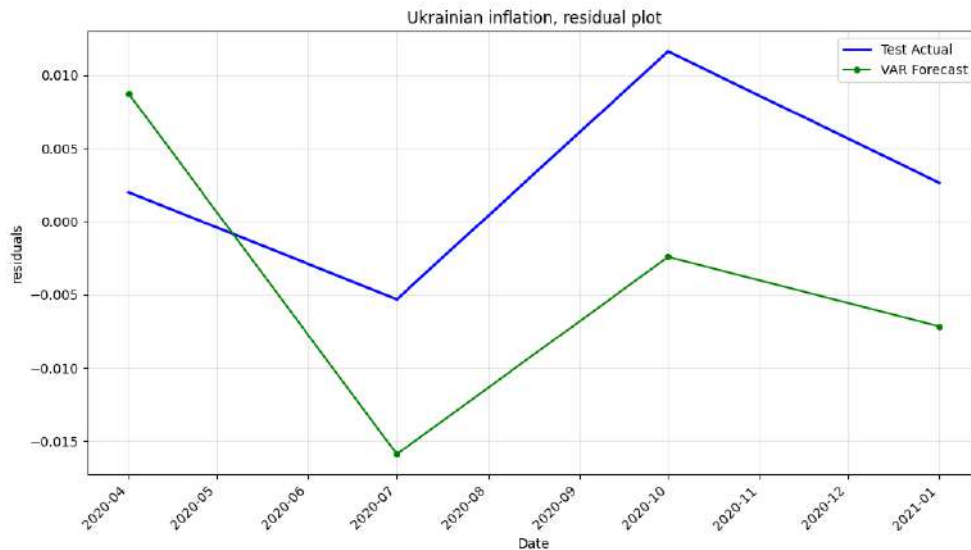


Рис. 3: Графік залишків для Інфляції України

Модель добре вгадала тренд для інфляції, що можна бачити з рисунку вище.

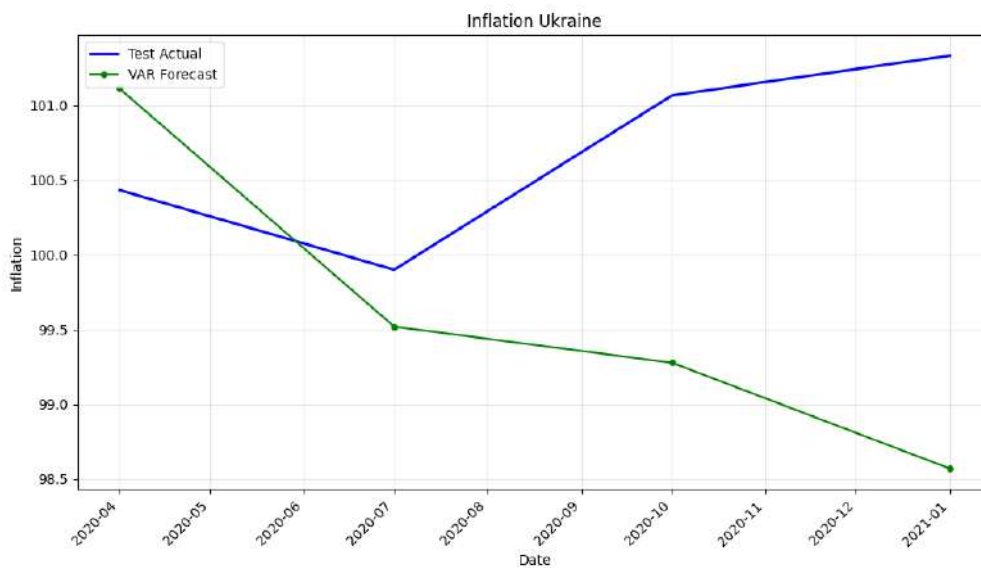


Рис. 4: Інфляція України

Для актуальних значень модель помилилась трохи з прогнозом на останній квартал 2020 року. Це пояснюється тим, що в кінці 2020 року вже почали знімати обмеження пов'язані з епідемією ковіду, тому і індекс цін зріс майже на 1%. Модель про це не знала, тому зробила неправильне припущення.

5.3 Імпульсно-перехідні функції (IRF)

В аналізі часових рядів імпульсно-перехідні функції дозволяють побачити як система реагує на “шок” однієї зі змінної. Вона відстежує взаємодинаміку змінної, у якої на попередньому кроці стався “шок” на те як це впливає на інші змінні. Зазвичай система швидко оговтується від раптових стрибків значень і вплив “шоків” згасає до нуля.

Наприклад, для ВВП України маємо такі графіки імпульсно-перехідних функцій:

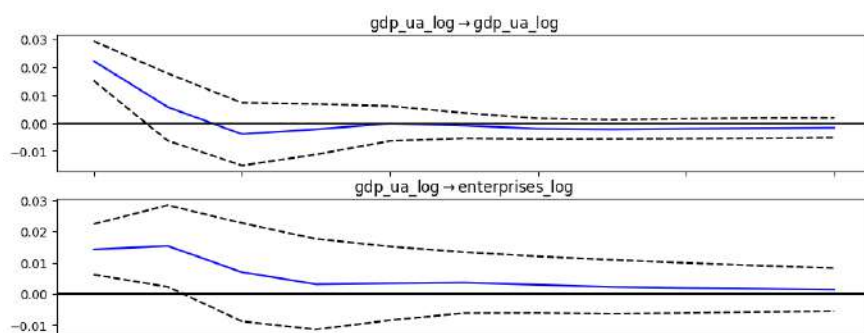


Рис. 5: Імпульсно-перехідні функції для ВВП

Як бачимо, є позитивний ріст ВВП та кількості підприємств. В перший період довірчі інтервали не включають нуль, тому можна сказати, що вплив “шоку” доволі суттєвий. З часом ці функції наближаються до нуля, що говорить про те, що вплив викиду є лише на найближчі квартали.

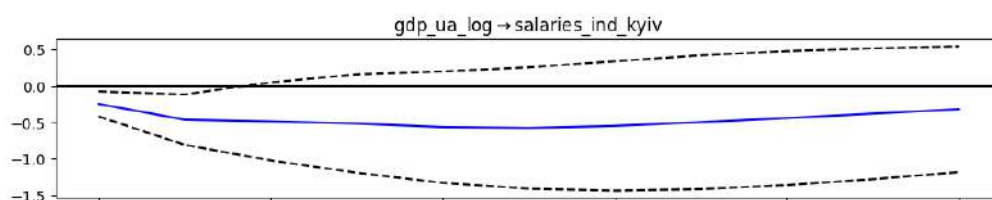


Рис. 6: Імпульсно-перехідні функції для ВВП

Єдиний негативний вплив при позитивному “шокові” є на індекс зарплат в Києві. Це можна пояснити тим, що при збільшенні ВВП покращується економічне життя регіонів і повільніше ростуть зарплати. Також може бути таке, що в модель не була включена важлива змінна, яка б могла пояснити це падіння.

Загалом динаміка для всіх змінних є схожою на динаміку графіків вище для ВВП. При позитивному збільшенні одного показника зазвичай ростуть і інші. Лише деякі показують спад, але це може бути з-за не доданих змінних до моделі.

6 Результати BVAR моделі

BVAR модель працює за схожим принципом як і попередня VAR модель, тільки вже вона складніша, адже в ній присутні Міннесотський пріор та ітеративний алгоритм Гіббса, які підганяють модель. Міннесотський пріор працює за принципом, що найкращий прогноз для змінної формується зі значення цієї ж змінної на попередньому кроці. Таким чином ми позбуваємось перенавчання моделі, де VAR модель намагалась підлаштуватись під кожну змінну. Це ж робить BVAR кращою для навчання на малих датасет, де перенавчання дуже легко отримати через малу вибірку даних. Алгоритм Гіббса підбирає ітеративним методом підбирає найкращі коефіцієнти для кожної змінної у моделі і бере медіанне значення. Таким чином ми отримаємо плавну лінію для прогнозів, адже на декількох тисячах ітерацій модель вирівнюється. Таким чином ми вирішуємо проблему шоків, коли прогноз може сильно піти вгору чи вниз. Ця ж модель припускає, що часовий ряд продовжить в майбутньому йти чіткою горизонтальною лінією, незважаючи на різкі шоки даних.

Цю модель було реалізовано з-за допомогою пакету BVAR[6] на мові програмування R[7]. BVAR модель також працює виключно з повними даними. Тому ми вимушені працювати з малим датасетом і генерувати відсутні квартальні дані з-за допомогою інтерполяції. Але завдяки методам Міннесотських пріорів та алгоритму Гіббса вона є стабільнішою і не зважає сильно на шоки.

6.1 Коефіцієнти для BVAR моделі

Як вже було сказано раніше BVAR модель працює за алгоритмами, які кажуть, що найкраще значення для змінної на наступному кроці це значення її на попередньому кроці. Модель зануляє майже всі інші параметри, адже так вона стає прогнозованішою та менш вразливою до шоків, наприклад, якщо одне зі значень сильно підніметься. Адже за економетричними припущеннями макроекономічні показники мають здатність відновлюватись від шоків, що лежить в основі BVAR моделі.

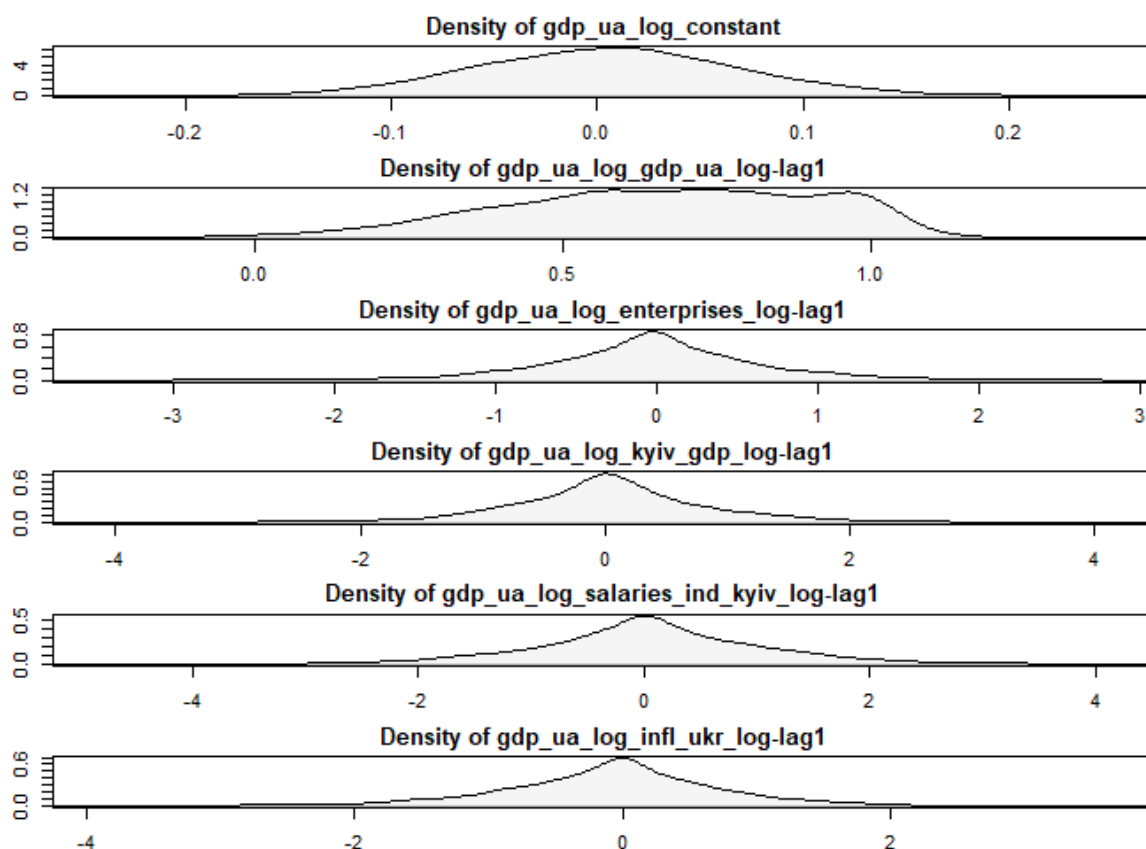


Рис. 7: Розподіл коефіцієнтів для ВВП України

На графіку ми бачимо розподіл коефіцієнтів, які визначила модель з-за допомогою алгоритму Гіббса з 5000 ітерацій. Можна бачити, що лише для ВВП України модель визначила ненульові коефіцієнти, що підтверджує роботу Міннесотського пріору, який вважає несуттєвими інші змінні у прогнозуванні. Але цей розподіл для ВВП України є доволі широким з чого можна сказати, що модель сумнівалась в виставленні коефіцієнтів, що скоріш за все свідчить на те, що є ще дані, які пояснюють ВВП України і не включені до цієї моделі.

Наступний графік ілюструє такий же розподіл коефіцієнтів, але вже для ВВП Київської області. І тут модель доволі впевнено визначила, що найбільше ВВП Київської області залежить від значень цієї ж колонки, але за попередній період. А всі інші змінні є несуттєвими. Також всі змінні мають доволі вузький дзвін розподілу, що говорить про те, що модель впевнена в коефіцієнтах.

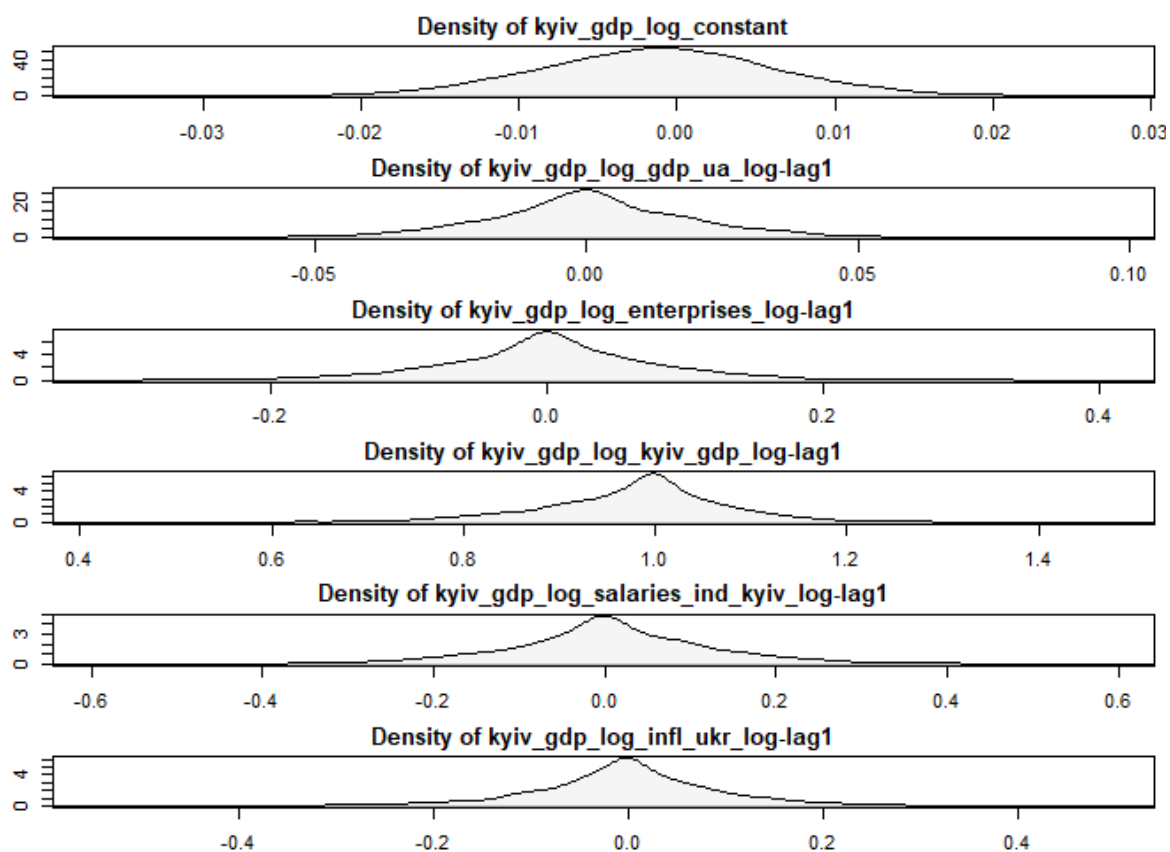


Рис. 8: Розподіл коефіцієнтів для ВРП Київської області

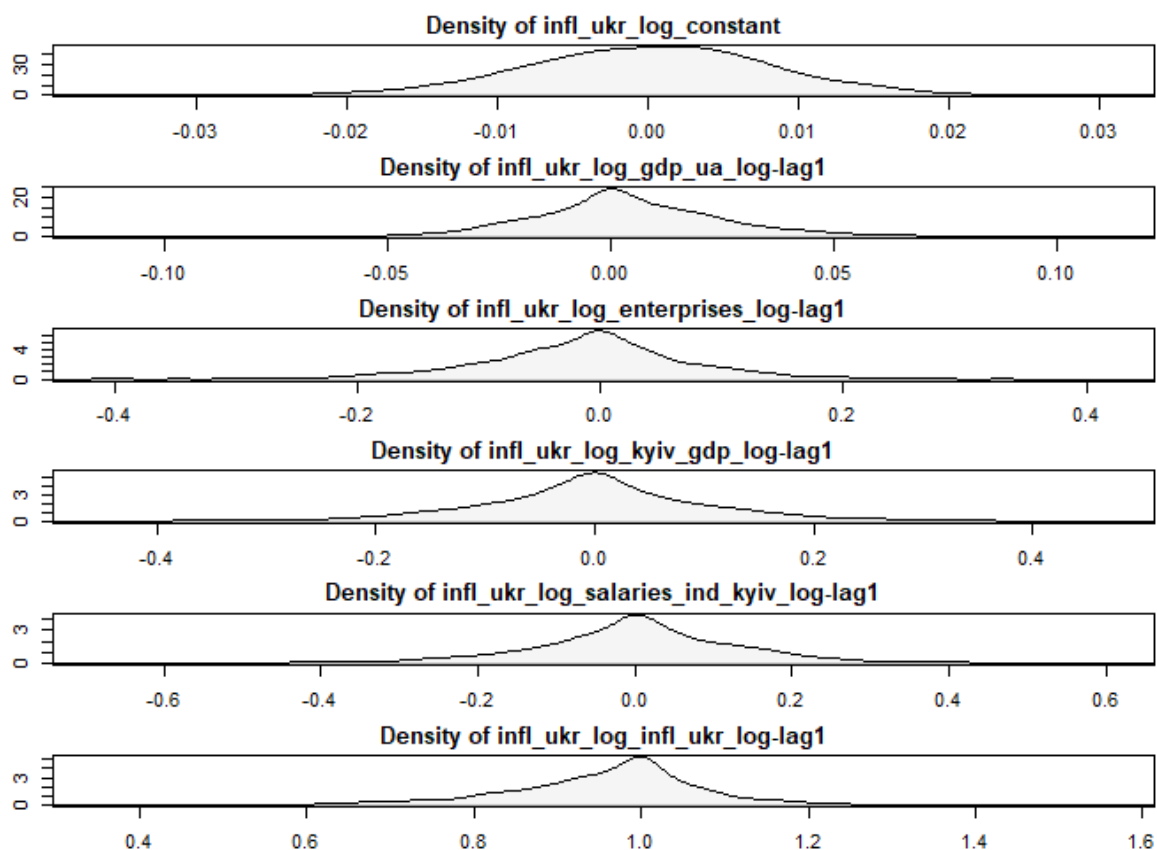


Рис. 9: Розподіл коефіцієнтів для інфляції України

Для інфляції України графіки розподілів такі самі як і для інших змінних. Модель тут також визначила, що найкраще наступне значення інфляції залежить від попереднього.

6.2 Прогнози BVAR моделі

Наступний графік демонструє прогнози BVAR моделі на тестову вибірку з 4 періодів. Він включає в собі прогнози для змінних: ВВП України, кількість підприємств, ВРП Київської області, зміна заробітної плати в Київській області та індекс цін в Україні.

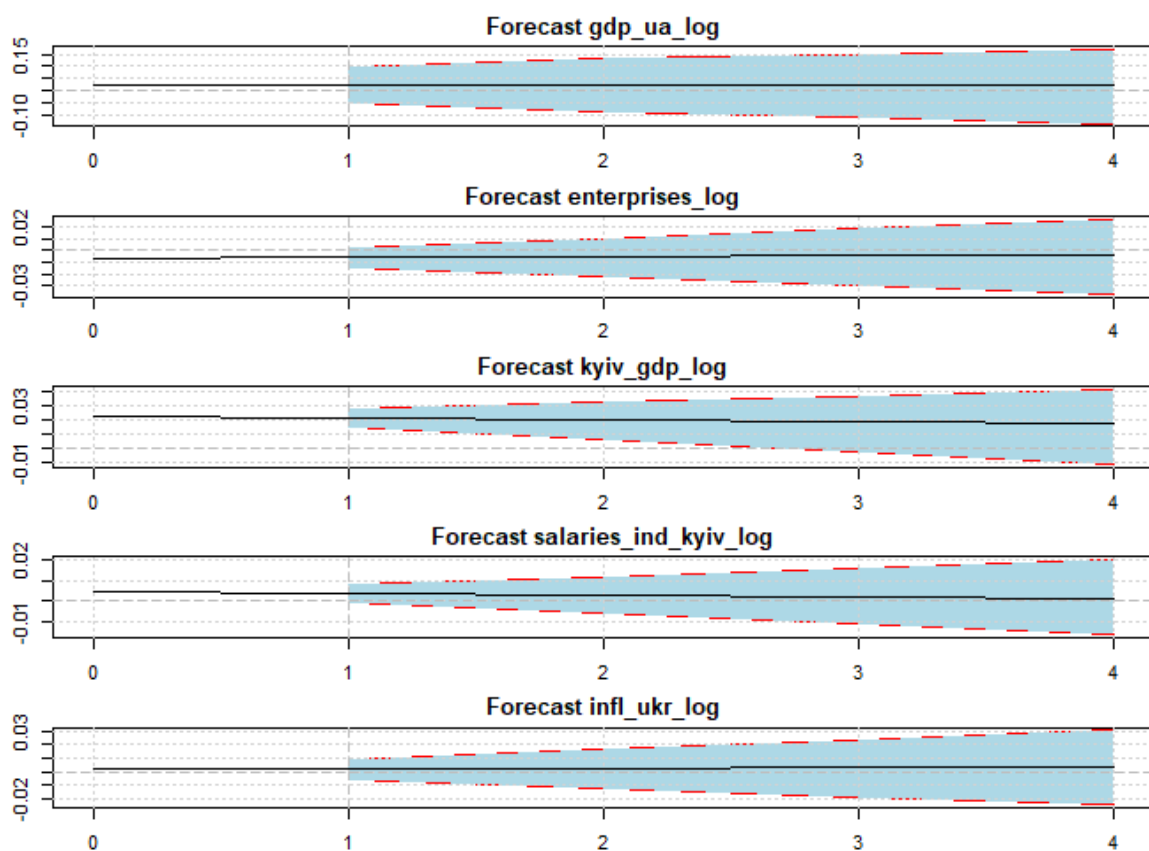


Рис. 10: Прогнози BVAR моделі

Як можна бачити, для всіх змінних модель прогнозує плавну динаміку розвитку на наступні 4 періоди. Все це з-за того, що BVAR модель зануляє не важливі змінні, щоб модель не перенавчалась. Саме тому прогнози ґрунтуються на тому, що найкраще значення для наступного періоду це значення на попередньому періоді цієї конкретної змінної плюс випадковий шум.

Якщо говорити більше про динаміку кожної конкретної змінної, то виді-

ляється ВВП України. ВВП дуже волатильна змінна, тому ми маємо такий широкий довірчий інтервал від -0.10 росту до 0.15. Серед інших змінних, які є не такими волатильними, довірчі інтервали доволі малі.

6.3 Імпульсно-перехідні функції

Як і для звичайної VAR моделі, для BVAR ми можемо побудувати Імпульсно-перехідні функції (IRF).

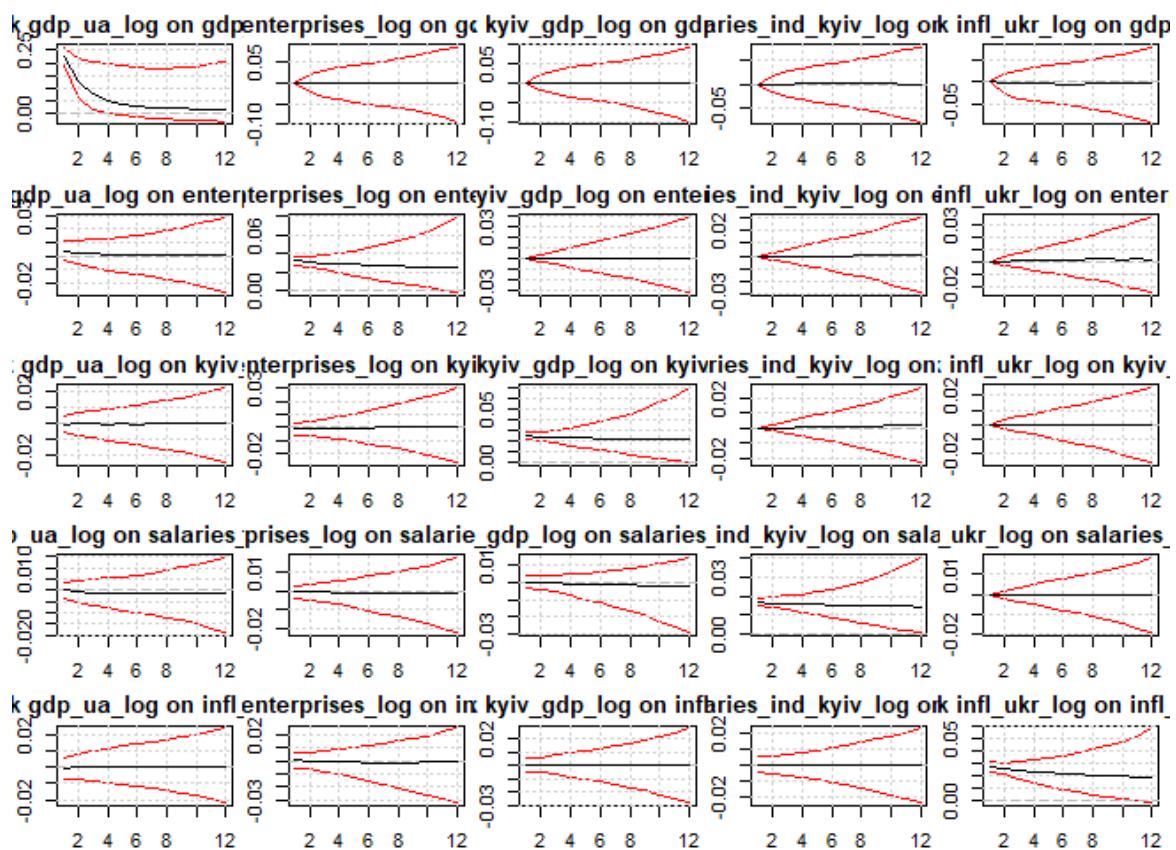


Рис. 11: IRF для BVAR моделі

Так як BVAR модель працює на пріорах, які зануляють інші змінні в прогнозі, то на графіку вище ми можемо бачити, що раптовий стрибок на одне стандартне відхилення впливає лише на ту змінну по якій був аномальний стрибок значень. Тобто раптова зміна у ВВП України впливає лише на ВВП України. Так само з інфляцією та іншими змінними. Найбільший вплив має ВВП України, який від неочікуваного стрибка вгору на наступний період підіймається на 0.25% приблизно. Всі інші змінні мають невисоку зміну для першого періоду в рамках від 0.01% до 0.05%. Також як можна бачити, до-

вірчі інтервали не перетинають нуль лише, коли змінна діє на саму себе, що знову впливає з пріорів, які були задані для моделі. З цього можна сказати, що ці зміни значущі, в інших нуль знаходиться в проміжку довірчих інтервалів, тому ми не можемо сказати, що ці зміни є значущими. На графіку вище показані раптові зміни на одне стандартне відхилення вгору. Імпульсно-перехідні функції симетричні відносно нуля, тому для раптової зміни на одне стандартне відхилення графіки будуть симетричними відносно порівняно з минулими графіками.

7 Результати моделі для змішано-частотних даних

У двох попередніх моделях два основних недоліки з-за яких ми не можемо отримати кращі прогнози. Перша проблема полягає в тому, що ми мали мало спостережень для тренування моделі і для деяких показників змушені були генерувати нові значення. Другою важливою проблемою є перенавчання моделей. Це стається з-за того, що ми намагаємось додати багато змінних, тому в попередніх моделях було використано лише декілька показників.

Ці проблеми вирішує модель *DynamicFactorMQ[3]* з пакету *statsmodels.api*[9]. Вона створена для того, щоб багато змінних помістити в один-два фактори, які будуть описувати ці дані. Таким чином ми тренуємо модель не на величезному датасеті, а на факторах, які описують лише наш датасет. Також важливою перевагою цієї моделі є те, що вона дозволяє працювати з пропущеними даними і робити прогноз на поточний період (nowcast).

7.1 Механіка роботи

Модель ґрунтується на наступній системі рівнянь:

$$\begin{cases} y_t = \Lambda f_t + \epsilon_t \\ f_t = A_1 f_{t-1} + \dots + A_p f_{t-p} + u_t \end{cases} \quad (1)$$

Перше рівняння $y_t = \Lambda f_t + \epsilon_t$ описує спостереження, які ми маємо. В цьому рівнянні y_t представляють реальні цифри по всім показникам, які є в нашій моделі. f_t це фактор, який ми хочемо оцінити, а Λ коефіцієнти для кожної змінної, які показують на скільки сильно вони впливають на фактор.

Друге рівняння представляє собою звичайну авторегресію AR(p), де ми хочемо описати фактор. Множники біля факторів означають матриці переходу, які показують на скільки сильно фактор на кроці t залежить від фактору на кроці $t - p$.

Для збіжності алгоритму використовують метод максимальної правдоподібності. А для знаходження проущених значень використовуємо фільтр Калмана, який за допомогою рівняння фактору знаходить взаємозв'язки і

дає оцінку невідомим значенням. Таким чином ми і робимо прогнозування на поточний період (nowcast) для змінних у яких ще немає даних.

7.2 Коефіцієнти моделі

Натренувавши модель, ми маємо наступні коефіцієнти зв'язку між факторами і всіма змінними:

Observation equation:

Factor loadings:	gdp_ua_log	infl_ukr_log	idiosyncratic: AR(1)	var.
infl_ukr_log	-0.01	-0.01	-0.06	0.98
gdp_ua_log	0.14	0.17	0.82	0.05
export_log	0.08	0.16	-0.58	0.31
import_log	0.06	0.21	-0.70	0.11
empl_ua_log	0.08	0.03	0.66	0.47
empl_kyiv_log	-0.06	-0.10	0.38	0.58
constr_ind_log	0.01	0.03	-0.43	0.80
house_ind_log	-0.02	0.07	-0.67	0.49
house_space_kyiv_log	-0.02	0.20	0.35	0.00
house_space_ua_log	-0.02	0.21	0.71	0.01
conf_ind_finance_log	0.15	-0.09	0.00	0.01
conf_ind_build_log	-0.01	0.04	0.70	0.50
conf_ind_service_log	0.12	-0.03	0.03	0.37
kyiv_gdp_log	0.01	0.17	-0.00	0.57

Рис. 12: Коефіцієнти зв'язку факторної моделі

Другий і третій стовпчик на рисунку показують наскільки сильно фактор впливає на незалежну змінну. Як можна бачити фактор ВВП найбільше впливає на макроекономічні показники як сам ВВП, експорт, імпорт, індекси ділової активності у фінансовій сфері та у сфері послуг. Загалом саме ці показники найбільше визначають ВВП, тому не дивно, що факторна модель показала такі коефіцієнти.

Третій стовпчик показує вплив фактора інфляції на інші змінні. І найбільший вплив діє на змінні ВВП, експорту, імпорту та показників житлової

площі по Україні та Київській області. Останні два показники дуже цікаві, адже можна зробити припущення, що ріст інфляції дуже впливає на ринок житла. Саме це стало чинником глобальної економічної кризи в 2008 році, коли позичальники коштів перестали платити відсотки по іпотеці з-за чого ринок житла впав. Тому це припущення цілком логічне щодо впливу інфляції на ринок житла України та Київської області.

Останній стовпчик показує скільки відсотків варіації пояснюється двома факторами. Для ВВП пояснюється 95% варіації, що дуже хороший результат. В той час як для інфляції дуже мало, що пов'язано з тим, що інфляція мало залежить від загальних факторів і залежить більше від значень інфляції на минулих кроках.

Також факторна модель дає змогу визначити скільки варіації для кожної змінної пояснює кожен фактор окремо.

	gdp_ua_log	infl_ukr_log
infl_ukr_log	0.031830	0.006996
gdp_ua_log	0.307867	0.443202
export_log	0.199027	0.452089
import_log	0.063011	0.678824
empl_ua_log	0.118931	0.026387
empl_kyiv_log	0.131075	0.129018
constr_ind_log	0.001205	0.064620
house_ind_log	0.021244	0.072640
house_space_kyiv_log	0.099720	0.997528
house_space_ua_log	0.136512	0.989354
conf_ind_finance_log	0.853659	0.265080
conf_ind_build_log	0.017319	0.016911
conf_ind_service_log	0.606133	0.053581
kyiv_gdp_log	0.151660	0.539318

Рис. 13: R_squared для кожної змінної

З малюнку видно, що найкраще перший фактор пояснює економічні індекси ділової активності у фінансовій сфері та сфері послуг і вони здебільшого

залежать від ВВП, тому і мають такий великий відсоток пояснення варіації. Для інфляції ситуація схожа, що і написана трохи вище, найвищі показники пояснення дисперсії для індексу цін на житло. Саме тому ми довели, що інфляція найпершим чином впливає на зменшення загальної площі житла через нездатність платити за житло.

7.3 Прогнози на майбутні періоди та поточне прогнозування

Специфіка цієї факторної моделі в тому, що окрім прогнозу майбутнє, ми можемо також прогнозувати пропущені значення за допомогою фільтру Калмана. Таким чином ми можемо прогнозувати як будуть рухатись квартальні показники, не маючи ще актуальних даних по цьому кварталу. Наступний графік ілюструє динаміку ВВП для прогнозів на майбутнє та по місяцям за минулий період.

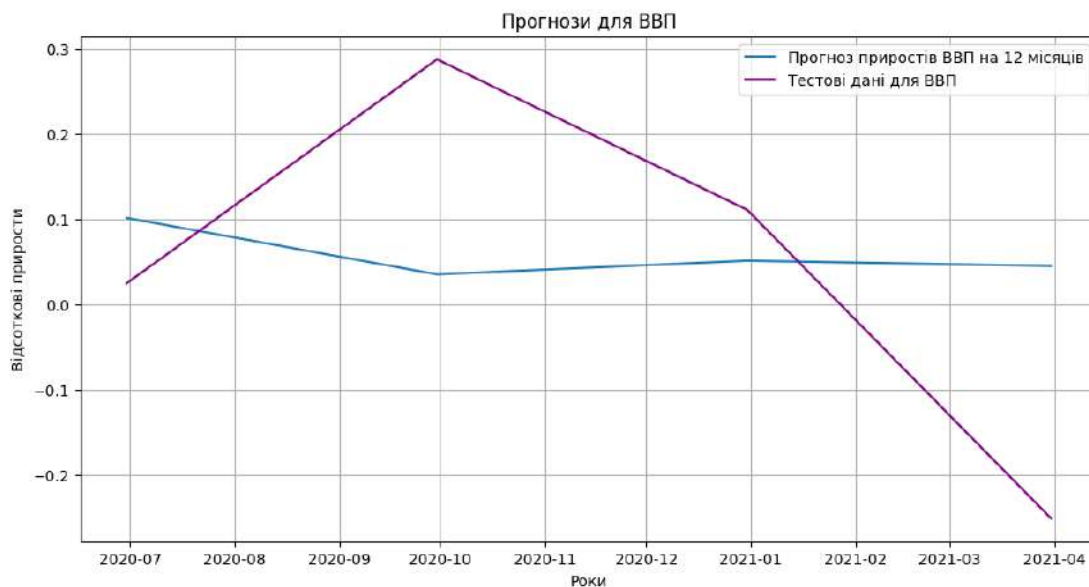


Рис. 14: Прогнозування для ВВП

Модель тренувалась на логарифмічних приростах, тому прогноз на рисунку теж у приростах. Факторна модель схожа на попередню BVAR, яка теж не ризикує у прогнозах. Завдяки тому, що ми працюємо зі стаціонарними даними, модель визначила, що найкращий прогноз на майбутнє це значення на попередньому кроці плюс шок. ВВП сильно волатильна змінна, саме тому

цій моделі було складно вгадати наступне значення показника.

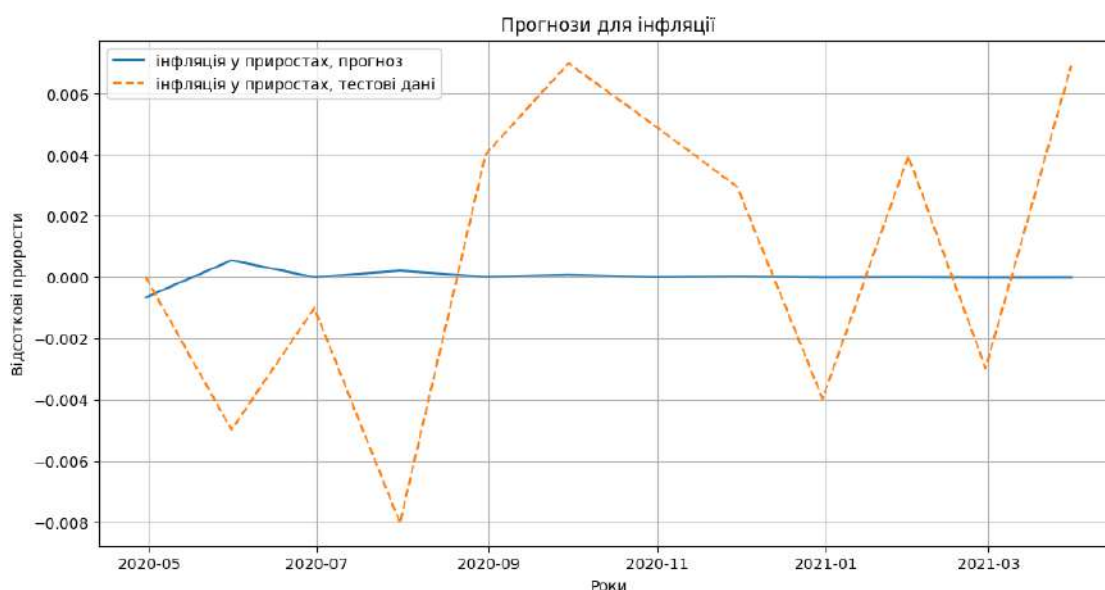


Рис. 15: Прогнозування для інфляції

Для інфляції з прогнозом на майбутній період схожа ситуація, адже модель визначила, що найкращий прогноз це рухатись далі майже прямою лінією. Але для інфляції ситуація краща, адже на відміну від ВВП, у інфляції немає сезонності та тренду. Тобто як можна бачити з графіку вище, оригінальний ряд вже є стаціонарним. Це покращує прогноз моделі, яка не ризикує і знає ряд приростів буде стрибати навколо прямої $y = 0$.

Тренуючи попередні моделі, було багато проблем зі змінними, адже вони мають квартальну періодичність, а іноді і річну взагалі, що не дозволяє робити якісні прогнози для моделей, адже ми стикаємось з пропущеними даними. Змішано-частотні моделі вирішують цю проблему методом Калмана чи схожими методами.

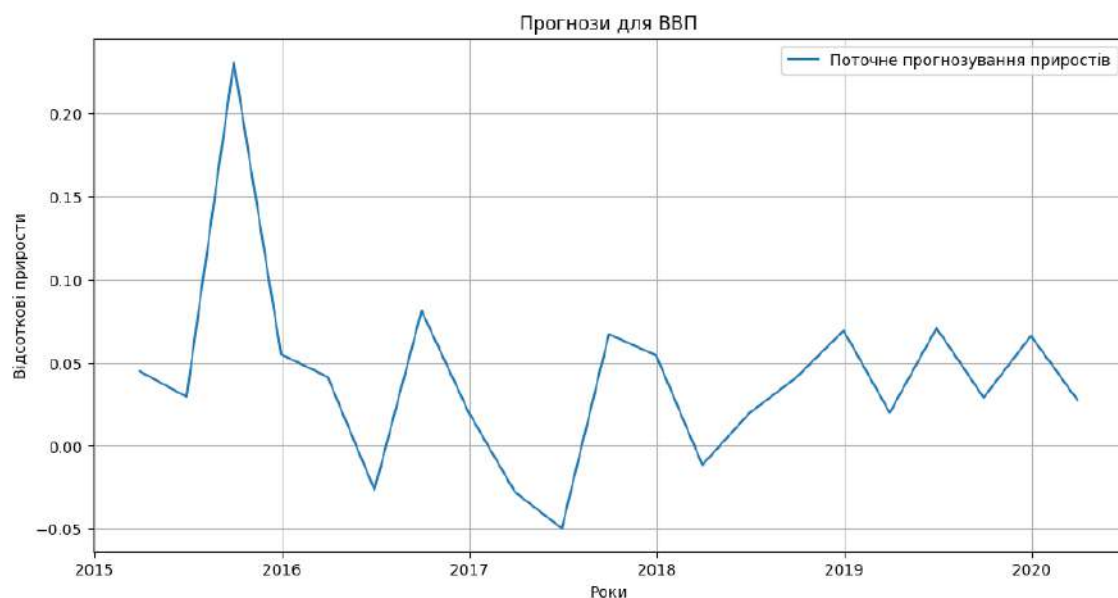


Рис. 16: Поточне прогнозування для ВВП

На графіку можна бачити результат поточного прогнозування для ВВП України. Основна риса цього прогнозування в тому, що ми можемо отримати динаміку даних ще до їх офіційної публікації. Таким чином ми можемо припустити, як рухатимуться дані і після публікації нових показників порівняти між собою.

8 Результати моделей

В процесі дослідження було натреновано 3 різних авторегресійних моделі: звичайну VAR модель, Байєсівську VAR модель та змішано-частотну факторну модель. Ці три моделі будемо порівнювати між собою на метриках RMSE та MAE, які визначають наскільки далекі прогнози від актуальних значень.

RMSE визначається як:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (Predicted_i - Actual_i)^2}{N}}$$

і позначає корінь з середньоквадратичної різниці між прогнозованими та реальними значеннями.

MAE визначається як:

$$MAE = \frac{1}{n} \sum_{i=1}^N |Predicted_i - Actual_i|$$

і вимірює середню величину помилки між прогнозованими та реальними значеннями. Різниця між цими методами, що RMSE краще вловлює викиди, бо сильно їх штрафує, тому такий показник корисний для волатильних змінних. MAE навпаки вираховує середню помилку, що робить цю метрику більш “чесною”.

	VAR	BVAR	DFM
RMSE	0.083	0.1953	0.1976
MAE	0.060	0.1585	0.1635

Табл. 2: RMSE і MAE для ВВП

На таблиці, яка розташована вище, можна бачити результати метрик для трьох моделей по ВВП України. Найкращі результати по RMSE показала проста VAR модель, це пов’язано з тим, що модель дуже добре завчила шоки та сезонність завдяки фіктивним змінним. BVAR і DFM мають набагато гірші показники по цій метриці. Все це з-за того, що ці моделі не зважають сильно на шоки від яких ВВП то зростає, то спадає. Вони будують стабільні моделі, які не залежать від неочікуваних майбутніх викидів, присутніх в

даних. Саме тому в розрахунку на довге майбутнє ці моделі є кращими для прогнозування.

По метриці MAE найкращою вийшла проста VAR модель. Ця метрика показала вже не такі великі відхилення, але всеодно з-за того, що VAR модель краще підлаштувалась під шоки та сезонність ВВП, саме ця модель є найкращою по двом метрикам.

	VAR	BVAR	DFM
RMSE	0.0106	0.00583	0.00486
MAE	0.0102	0.00445	0.00433

Табл. 3: RMSE і MAE для інфляції

Метрики RMSE та MAE для інфляції показали вже зовсім іншу динаміку, що була для ВВП на таблиці вище. Тепер найкращими моделями є BVAR та DFM через те, що ці моделі менше зважають на шоки інших змінних і будують прогноз зважаючи на значення цільової змінної за минулі проміжки часу. А інфляція загалом залежить лише від власних значень на попередніх періодах, тому і прогноз вийшов таким точним. VAR модель, яка намагалась підлаштуватись під шоки тут виявилась найгіршою.

9 Висновки

В ході цієї роботи було натреновано 3 різні авторегресійні моделі. Результатом цих тренувань стали прогнози нових значень для макроекономічних показників України та Київської області. В процесі дослідження виникало багато проблем з налаштуванням моделей, адже багато макроекономічних даних публікуються лише щорічно, тому вибірка дуже мала для тренування і з-за цього результати можливо не такі хороші як хотілось би. Але попри всі труднощі вдалось отримати хороші показники прогнозованих значень, також окрім прогнозування на майбутні періоди, завдяки моделі DFM вдалось отримати значення для поточного прогнозування змінних (nowcast). Тому завдяки цій моделі вийшло знайти значення ВРП Київської області по місяцям, коли цей показник публікується лише щорічно.

В підсумку, можна сказати, що змішано-частотні авторегресійні моделі є ефективнішими для прогнозування макроекономічних показників, адже для цих моделей не треба брати значення виключно по кварталам чи генерувати пропущені змінні, бо завдяки фільтру Калмана вони знаходять пропущені дані для показників. Саме тому використання змішано-частотних авторегресійних моделей є кращим для прогнозування макроекономічних показників.

10 Додатки

Дані та весь код для моделей машинного навчання, які були натреновані для цієї роботи, знаходяться за посиланням:

https://github.com/mixoill/varmodels_Siryi

Литература

- [1] Bhandari Anmol, Argente David, Lee Munseob, and Moreira Diana. The cost of privacy: Welfare effects of the disclosure of personal data. Working Paper 787, Federal Reserve Bank of Minneapolis, 2026. Accessed: 24.02.2026. URL: <https://researchdatabase.minneapolisfed.org/downloads/rx913p93b>.
- [2] Joshua CC Chan. Large bayesian vars with adaptive hierarchical priors. *Journal of Econometrics*, 221(1):115–130, 2021. URL: <https://joshuachan.org/papers/BVAR-MAHP.pdf>.
- [3] Chad Fulton. Large dynamic factor models, 2024. Accessed: 04.05.2026. URL: <https://www.chadfulton.com/topics/state-space-large-dynamic-factor-models.html>.
- [4] Koop Gary, McIntyre Shawn, and Mitchell James. Incorporating short data into large mixed-frequency vars for regional nowcasting. *Journal of the Royal Statistical Society: Series A*, 2023. doi:10.1111/rssa.13050.
- [5] GeeksforGeeks. Vector autoregression (var) for multivariate time series. <https://www.geeksforgeeks.org/machine-learning/vector-autoregression-var-for-multivariate-time-series/>, 2026. Accessed: 20.02.2026.
- [6] Kevin Kotze. Bayesian vector autoregression models. Accessed: 04.05.2026. URL: <https://kevinkotze.github.io/ts-9-tut/>.
- [7] Nikolas Kuschnig and Lukas Vashold. *BVAR: Bayesian Vector Autoregressions*. Comprehensive R Archive Network (CRAN), 2021. R package version 1.0.2. URL: <https://cran.r-project.org/web/packages/BVAR/BVAR.pdf>.
- [8] Christopher A. Sims. Macroeconomics and reality. *Econometrica*, 48(1):1–48, 1980. doi:10.2307/1912017.
- [9] Statsmodels Developers. statsmodels.tsa.statespace.dynamic_factor.dynamicfactor. <https://www.statsmodels.org/dev/generated/statsmodels.tsa>.

`statespace.dynamic_factor.DynamicFactor.html`, 2024. Accessed: 2026-05-13.

- [10] Statsmodels Developers. Vector autoregressions (var) - statsmodels 0.15.0.dev0 documentation. https://www.statsmodels.org/dev/vector_ar.html, 2024. Accessed: 2026-05-13.
- [11] Вікіпедія. Баєсове висновування. https://uk.wikipedia.org/wiki/Bayesian_inference, 2026. Accessed: 20.02.2026.
- [12] Державна служба статистики України. Статистична інформація. <https://stat.gov.ua/uk>, 2025. Accessed: 16.12.2025.