Застосування динамічних факторних моделей в задачах прогнозування макроекономічних показників

Магістерська дисертація

студентки 6 курсу групи КА-41м

ННК “ІПСА” НТУУ “КПІ”

Урупи Анастасії Віталіївни

Науковий керівник:
канд. фіз.-мат. наук, доцент
Каніовська І.Ю.

Вступ

- Економічні дані характеризуються складними внутрішніми зв’язками між параметрами, які важко виразити функціонально, і одним із методів моделювання таких даних є використання факторних моделей.
- Застосування факторних моделей
  - Побудова економічних показників(Chicago Fed National Activity Index (CF-NAI), EuroCoin)
  - Прогнозування
  - Аналіз валютної політики.
  - Міжнародні бізнес-цикли

Мета дослідження

- **Об’єкт дослідження:** моделювання макроекономічних показників.
- **Предмет дослідження:** методи оцінки кількості факторів та оптимального лагу у динамічній факторній моделі.
- **Мета дослідження**:
  - Розглянути різні методи оцінки параметрів динамічної факторної моделі
  - Порівняти поведінку цих методів на реальних даних за депозитними банківськими ставками України
  - Розробити рекомендації для вибору та використання цих методів

Факторна модель

- **Ідея:**

Описати варіацію спостережуваних змінних через декілька неспостережуваних, кількість яких зазвичай значно менша. Таким чином, припускається, що зміни у значеннях великої кількості показників спричинені деяким обмеженим набором змінних, які називаються *факторами*.

- **Формалізація:**

Спостережувані змінні моделюються як сума лінійних комбінацій факторів та випадкової змінної («похибки»).

Динамічна факторна модель

Нехай x_t – це N-вимірний вектор спостережуваних даних в моменти часу t, який визначається q динамічними факторами u_t з лагом до p, та власною незалежною компонентою ε_t за співвідношенням:

$$x_t = B_0 u_t + B_1 u_{t-1} + \ldots + B_p u_{t-p} + \varepsilon_t$$

В цьому рівнянні матриці B_i називаються матрицями *факторних навантажень* і визначають ступінь залежності спостережуваної змінної від факторів.

Модель у просторі стану:

$$F_t = AF_{t-1} + Gw_t$$

$$x_t = BF_t + \varepsilon_t$$

Визначення оптимальних кількості факторів та лагу на основі інформації в автоковаріаційних матрицях

Для застосування методу для факторної моделі робляться наступні припущення:

- q-вимірний вектор факторів u_t є Гауссівським білим шумом, тобто $Eu_t=0$ і $Du_t=I_q$;
- власні компоненти ε_t є Гауссівськими білими шумами з $E\varepsilon_t=0$ і $D\varepsilon_t=V$;
- фактори u_t та власні компоненти ε_t є незалежними між собою;
- матриця навантажень B має повний ранг, тобто rank(B)=(p+1)q=k .

Інформація Кульбака - Лейблера

Інформація (відстань) Кульбака - Лейблера – міра інформації, втраченої при заміні одного ймовірнісного розподілу іншим.

$$I(f,g) = E_f\left(\ln\frac{f(x)}{g(x)}\right)$$

Для матриці Γ_0: *f(x)* – Гауссівська щільність розподілу з нульовим мат. сподіванням та дисперсією $\Gamma_0 = C\Lambda C^T,\ \Lambda = diag(\lambda_1,\dots,\lambda_N)$ *g(x)* – Гауссівська щільність розподілу з нульовим мат. сподіванням та дисперсією $\Gamma_0 = I_N$ $I(f,g) = \sum_{j=1}^{N}\frac{(1-\lambda_j^2)}{\lambda_j}$	**Для матриці Γ_i, $i>0$** *f(x)* – умовна Гауссівська щільність y_t за відомого y_{t-i} з мат.сподіванням $\Sigma_i y_{t-i}$ та дисперсією $(I_N - \Sigma_i^2)$, де $\Gamma_0^{-1/2}\Gamma_i(\Gamma_0^{-1/2})' = H_i\Sigma_i Q_i$ *g(x)* – Гауссівська щільність з нульовим мат. сподіванням та одиничною дисперсійною матрицею. $I(f,g) = -\sum_{j=1}^{N}\ln(1-\rho_{i,j}^2)$

Критерій Бартлета

Таким чином, інформація визначається сингулярними значеннями розкладу матриці:

$$\Gamma_0^{-1/2}\Gamma_\iota(\Gamma_0^{-1/2})^T = H_\iota \Sigma_\iota \Theta_\iota$$

$$I(f,g) = -\sum_{j=1}^{N} \ln(1-\rho_{i,j}^2)$$

<u>Зокрема, якщо ці значення є нульовими, то інформація дорівнює нулю.</u>

Для перевірки гіпотези про те, що *N-k* найменших сингулярних значень рівні нулю, застосовуємо критерій Бартлета:

$$\chi^2 = -[T - \frac{1}{2}(2N+1)] \sum_{j=k+1}^{N} \ln(1-\rho_{i,j}^2)$$

Статистика критерію розподілена за законом χ^2 з *(N-k)*^{2} ступенями вільності.

Визначення оптимальної кількості факторів та лагу

Ранги матриць оцінюються за числом статистично значущих сингулярних значень за допомогою критерію Бартлета наступним чином. Якщо для заданого рівня значущості гіпотеза про те, що:

1. всі сингулярні значення для *(p+1)*-ої автоковаріації дорівнюють нулю, приймається
2. всі сингулярні значення для (p)-ої автоковаріації дорівнюють нулю, відхиляється
3. *(N-q)* найменших сингулярних значень дорівнюють нулю, приймається

Тоді *p є оцінкою порядку лагу*, а *q є оцінкою числа факторів*, що входять у модель.

Визначення оптимальних кількості факторів та лагу на основі похибки передбачення

1. Застосуємо сингулярний розклад до матриці X:

$$X = USC^T$$

де S – діагональна матриця, U та C – унітарні матриці.

2. Для фіксованого k застосуємо наступний розклад:

$$X = U_k S_k C_k^T + U_2 S_2 C_2^T = \hat{X} + E$$

де $U = [U_k \mid U_2], C = [C_k \mid C_2], S = [S_k \mid S_2].$

3. Транспонуючи цей розклад, отримуємо:

$$X^T = \hat{X}^T + E^T,$$

$$\hat{X}^T = C_k S_k [\hat{F}_1 \quad \dots \quad \hat{F}_T],$$

4. Таким чином, вектори $\hat{F}_1, \dots, \hat{F}_T$ приймаються за оцінки факторів моделі, а матриця $\hat{B} = C_k S_k$ є оцінкою навантажень цих факторів.

Визначення оптимальних кількості факторів та лагу на основі похибки передбачення

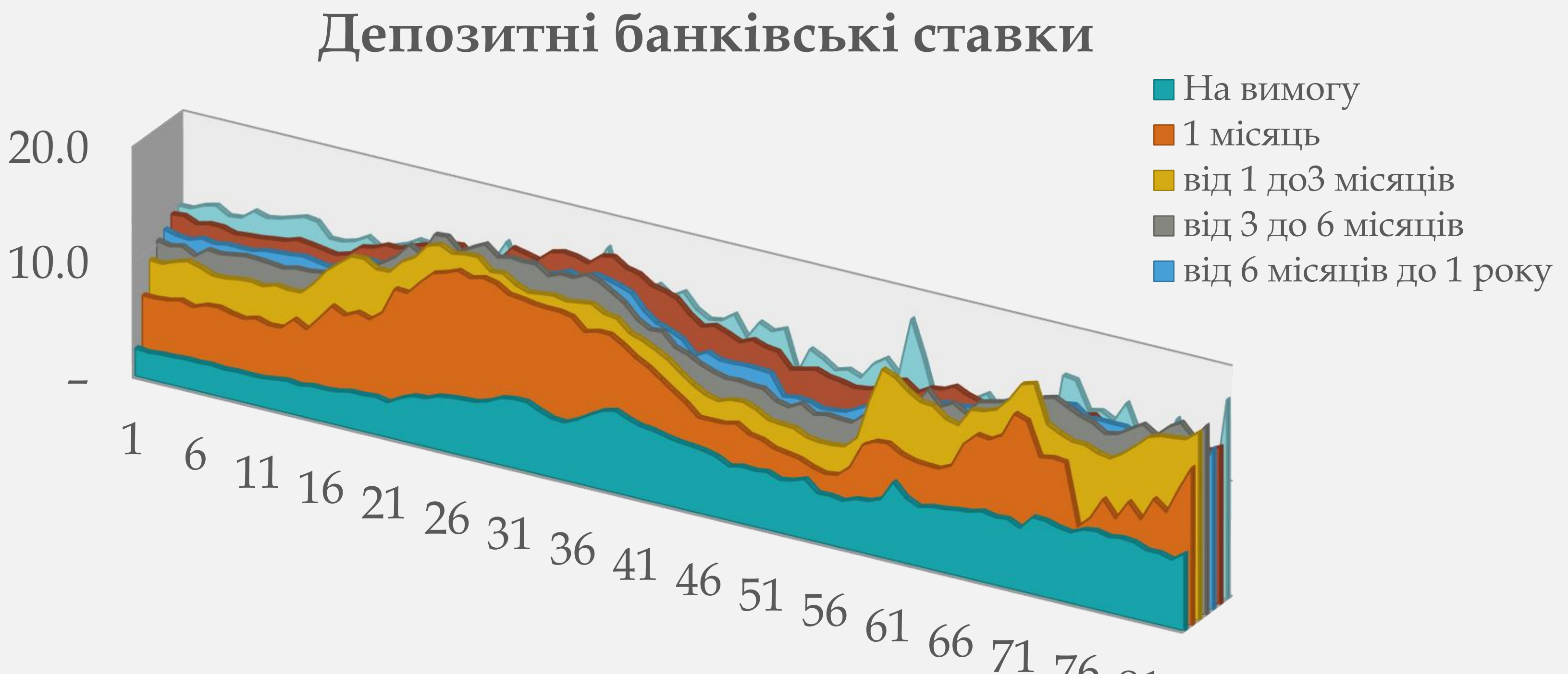
- Параметри моделі оцінюються не за усіма наявними спостереженнями, а за деяким числом спостережень T_1, а потім починає свою роботу фільтр Калмана. Тоді квадрат похибки передбачення:

$$\|PE\|_E^2 = \sum_{i=1}^{N} \sum_{t=T_1+1}^{T} (pe_{it})^2 = tr(PE^T PE)$$

- За такого підходу, число факторів, що включається в модель, визначається на основі <u>мінімізації похибки передбачення</u> при найменшому можливому числі факторів.

Дані для практичної задачі

- Поведінка методів була проаналізована при використанні на реальних даних.
- В якості вхідних даних були взяті щомісячні дані за депозитними банківськими ставками Національного банку України за останні 10 років. Дані складаються з 7 змінних і 86 спостережень.



Демонстрація методу на тестових даних

- Результати роботи першого методу на основі інформації в автоковаріаціях на тестових даних, які були згенеровані за факторною моделлю з параметрами *p=3, q=3, N=10, T=100.*

Parameters	Test statistic	P-value
p=1, q=0	973.247	0.0000
p=2, q=0	559.1482	0.0000
p=3, q=0	276.9956	0.0000
p=4, q=0	116.9822	0.1179
p=3, q=1	160.8497	3.233605e-07
p=3, q=2	95.9366	0.0059
p=3, q=3	37.5859	0.8826

- Отже, за результатами роботи методу оптимальні параметри моделі: p=3, q=3, що співпадає з початковою моделлю.

Практична задача

- Результати роботи першого методу на основі інформації в автоковаріаціях на даних за депозитними ставками України:

Parameters	Test statistic	P-value
p=1, q=0	754.7289	0.0000
p=2, q=0	583.3474	0.0000
p=3, q=0	513.0338	0.0000
p=4, q=0	474.7252	0.0000
p=5, q=0	467.0947	0.0000
p=6, q=0	443.0191	0.0000
p=7, q=0	435.2781	0.0000

- В цьому випадку метод не дав бажаних результатів, оскільки дані є сильно корельованими навіть на віддалених один від одного вимірах.

Результати роботи методу на основі похибки передбачення

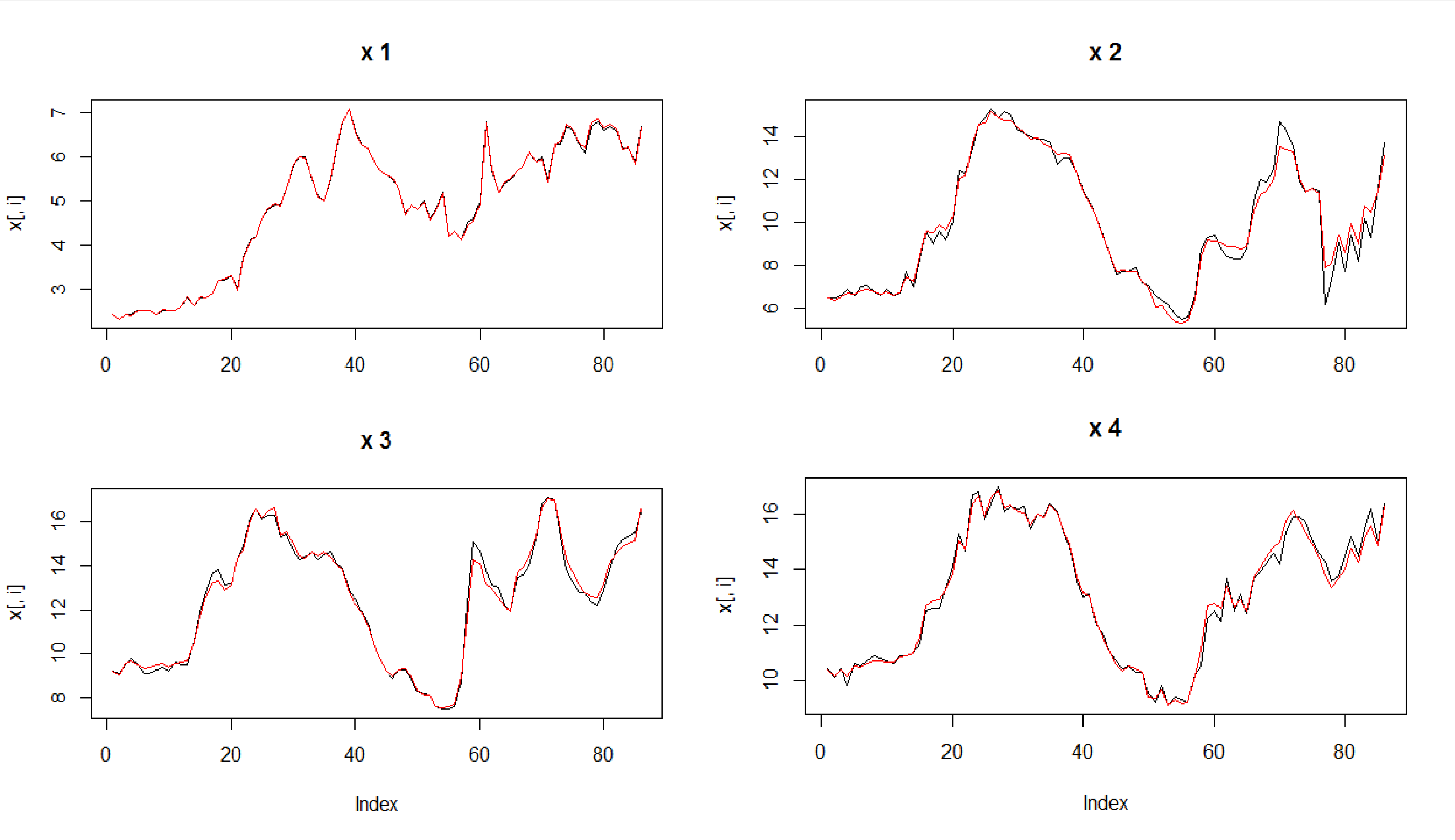
Параметри моделі оцінювалися на основі 50 спостережень, на основі даних, що залишилися, була оцінена похибка передбачення за допомогою фільтра Калмана.

Похибки передбачення для різної кількості факторів:

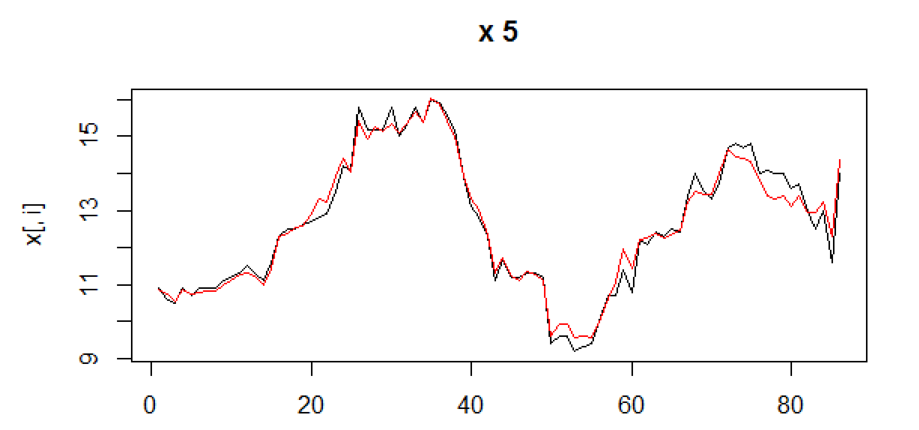
k	1	2	3	4	5	6
PE	0.1839	0.0510	0.0122	0.0071	0.0068	0.0067

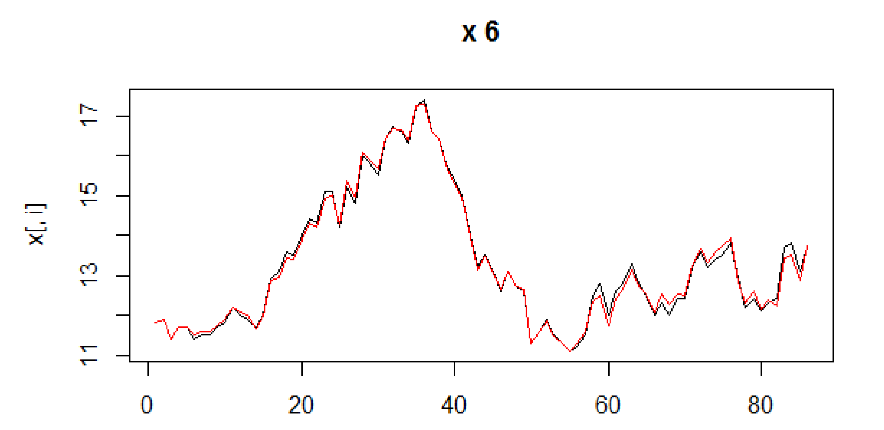
Як бачимо, 4 є оптимальним числом факторів для моделі, оскільки подальше збільшення числа факторів не призводить до суттєвого зменшення похибки передбачення.

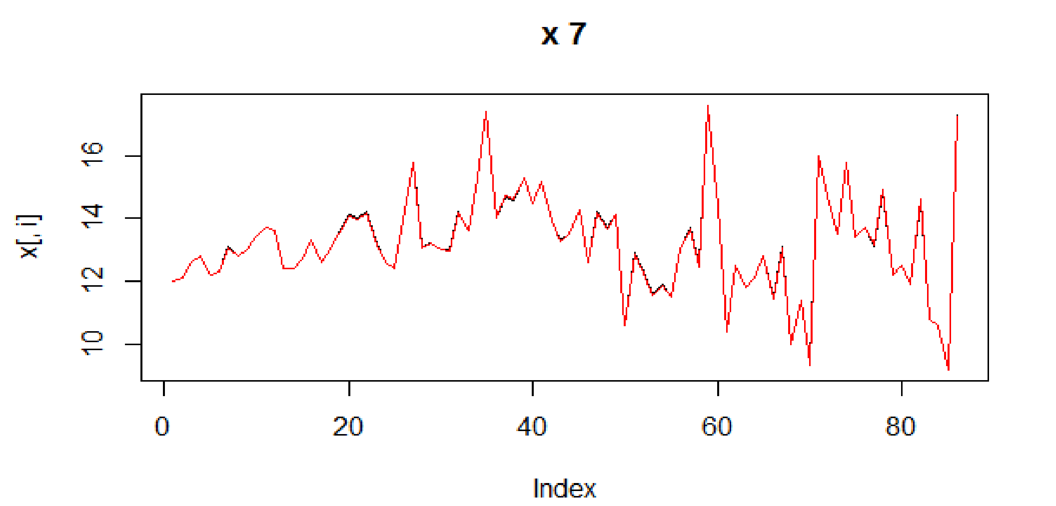
Графічне зображення результатів моделювання



Графічне зображення результатів моделювання







Порівняння методів на основі практичних результатів

Метод на основі інформації в автоковаріаційних матрицях:

+ є статистичним, тобто його достовірність залежить від кількості даних;
+ менш обчислювально затратний;
+ дає однозначний результат;
✕ незастосовний до даних, що сильно корельовані на різних вимірах.

Метод на основі мінімізації похибки передбачення:

+ більш наочний;
+ універсальний, тобто підходить для будь-яких даних;
✕ більш обчислювально затратний;
✕ результати можуть бути неоднозначними.

Висновки

- ✓ Розглянута ідея та формалізація динамічної факторної моделі.
- ✓ Розглянуті два методи вибору оптимальних кількості факторів та максимального лагу в моделі: на основі інформації, що міститься в автоковаріації, та на основі похибки передбачення.
- ✓ Методи застосовані на реальних даних за депозитними банківськими ставками України.
- ✓ На основі практичної задачі проведений порівняльний аналіз методів і надані рекомендації щодо вибору методів в залежності від вхідних даних.
- ✓ Новизна роботи полягає в застосуванні альтернативного методу, а саме методу, що заснований на інформації в автоковаріаційних матрицях, до моделювання макроекономічних даних, та порівнянні його зі стандартним методом на основі похибки передбачення.

Дякую за увагу!