



Національний технічний університет України “КПІ”
ІНН “Інститут прикладного системного аналізу”

Магістерська дисертація

тема:

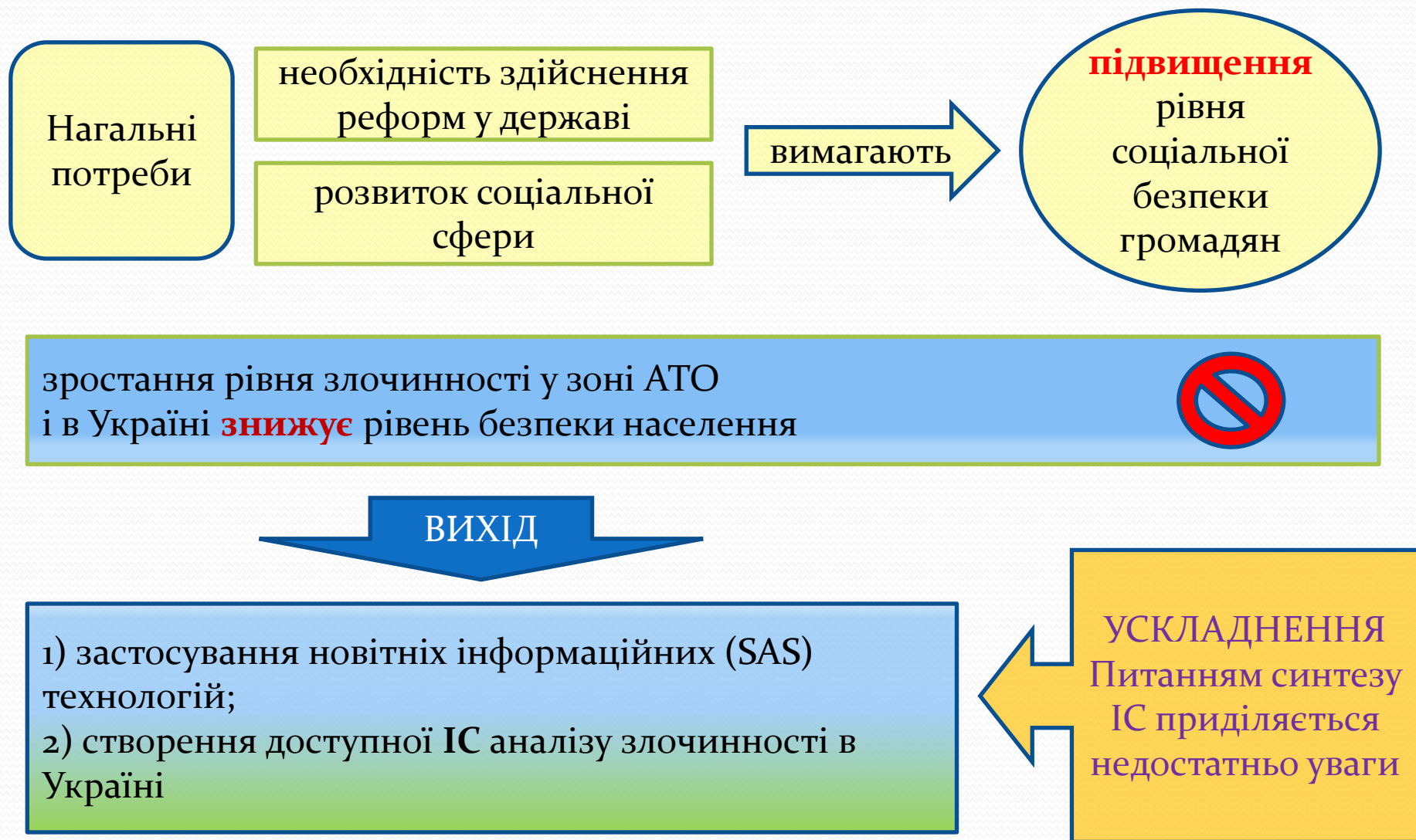
*“Прогнозування часових рядів із використанням
методу подібних траєкторій”*

Виконав: студент гр. КА-53,м
Щука Роман Володимирович

Київ 2017

Науковий керівник: к.т.н., м.н.с.
асистент каф. ММСА ІНН ІПСА О.М. Терентьєв

Актуальність теми дослідження



Об'єкт, предмет та мета дослідження

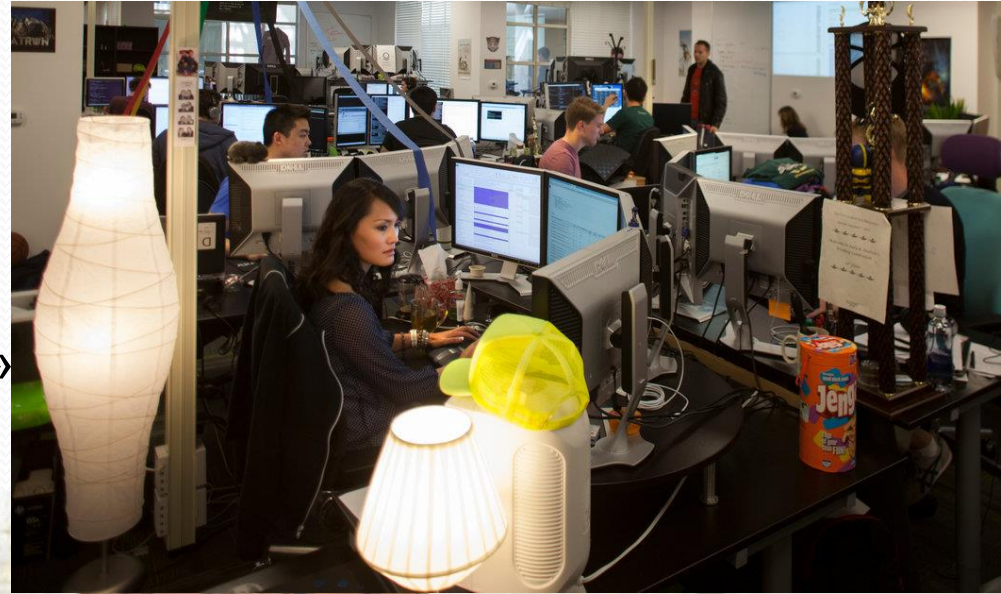
- ❑ **Об'єктом** дослідження є боротьба зі злочинністю в Україні.
- ❑ **Предметом** дослідження є прогнозування рівня злочинності в Україні з використанням методів аналізу часових рядів.
- ❑ **Мета** роботи – розробити алгоритм для прогнозування рівня злочинності в Україні та довести його до програмної реалізації на платформі SAS.

Завдання дослідження

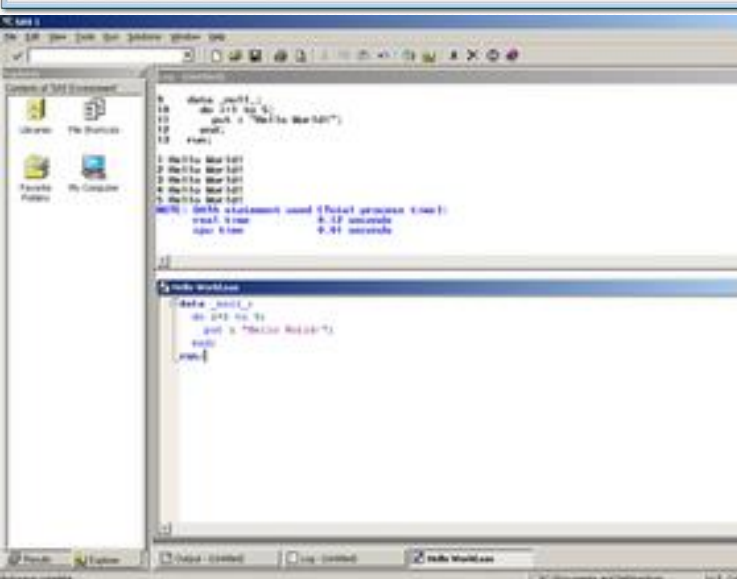
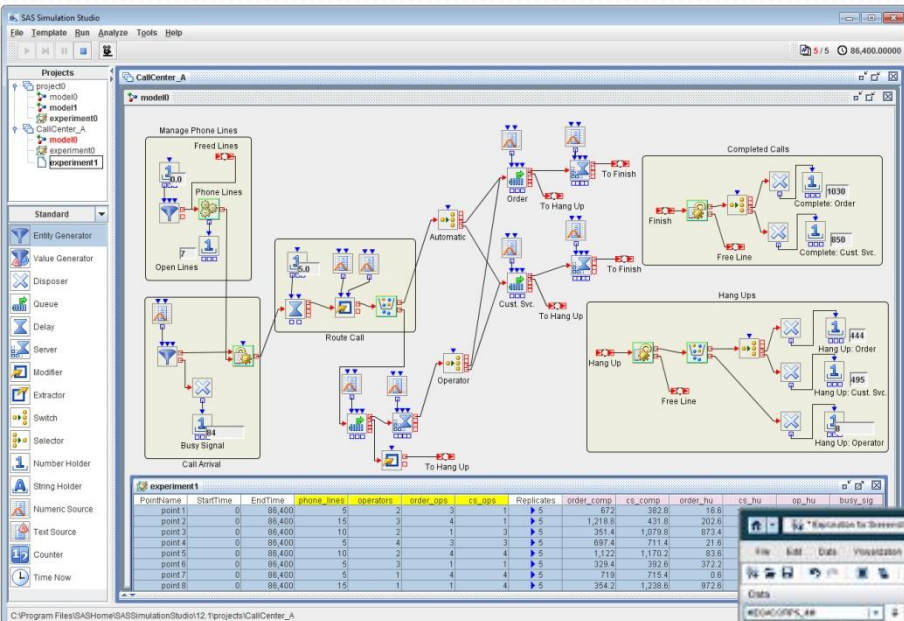
- ☐ аналіз предметної галузі та існуючих ІАС
- ☐ дослідження методів аналізу часових рядів і математичних моделей;
- ☐ синтез алгоритму прогнозування злочинності на основі методу подібних траєкторій;
- ☐ програмна реалізація на платформі SAS Enterprise Guide синтезованого алгоритму;
- ☐ розробка стартап проекту, його ринкової стратегії та маркетингової програми;

Передові інформаційні технології в галузі боротьби зі злочинністю

- ☐ Palantir
- ☐ IC «Контртерористичний центр»
- ☐ IC «Ситуаційний центр РНБО»
- ☐ SAS Memex
- ☐ SAS Fusion Center



4omy SAS?



Огляд SAS Software

Для моделювання часових рядів система SAS забезпечена процедурами:

- економетричного аналізу
- аналізу часових рядів
- прогнозування часових рядів
- моделювання та симуляції
- аналізу дискретного вибору
- аналізу залежних змінних моделі
- аналізу сезонності часових рядів даних
- фінансового аналізу і звітності
- доступу до економічних і фінансових баз даних
- управління даними часового ряду

В системі SAS запропоновані наступні методи аналізу, моделювання та прогнозування часових рядів:

- Регресійний аналіз:
 - лінійна регресія
 - нелінійна регресія
 - логістична регресія
- Кореляційний аналіз
- Факторний аналіз
- Кластерний аналіз
- Дискримінантний аналіз
- SAS Basic Forecasting:
 - покрокова авторегресія
 - експоненційне згладжування
 - модель Уінтерса
- інші методи

Перелік основних методів та моделей прогнозування

Методи адаптивного прогнозування: Методи регресійного аналізу:

- експоненційне згладжування;
- моделі лінійного росту:
- Хольта;
- Брауна;
- Тейла-Вейджа;
- Бокса-Дженкінса;
- сезонні моделі Вінтерса.

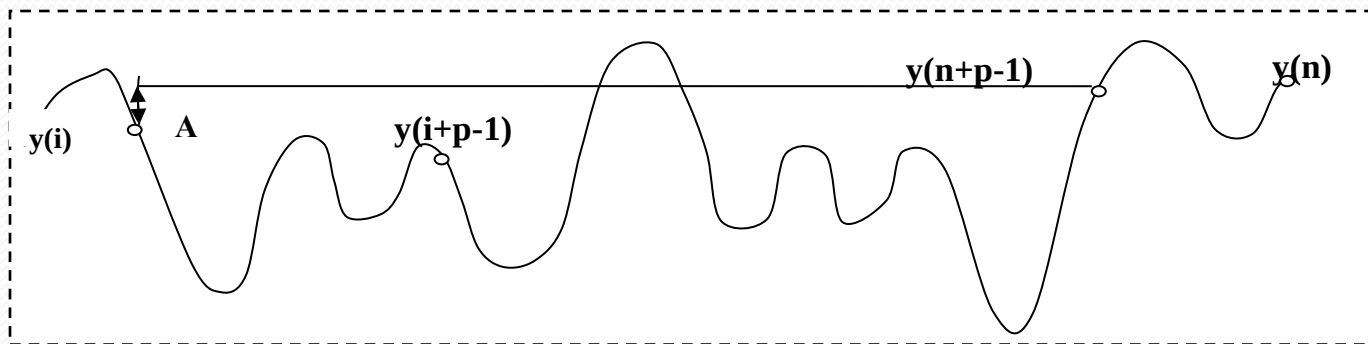
- АР – моделі з авторегресією складовою;
- АРКС – модель авторегресії із ковзним середнім
- середнім;
- моделі у вигляді тренду;
- АРІКС моделі;
- ARMAX –моделі.
- множинна регресія.

Методи класифікації:

- Метод подібних траєкторій (МПТ);
- Кластерний аналіз;
- Машина опорних векторів.

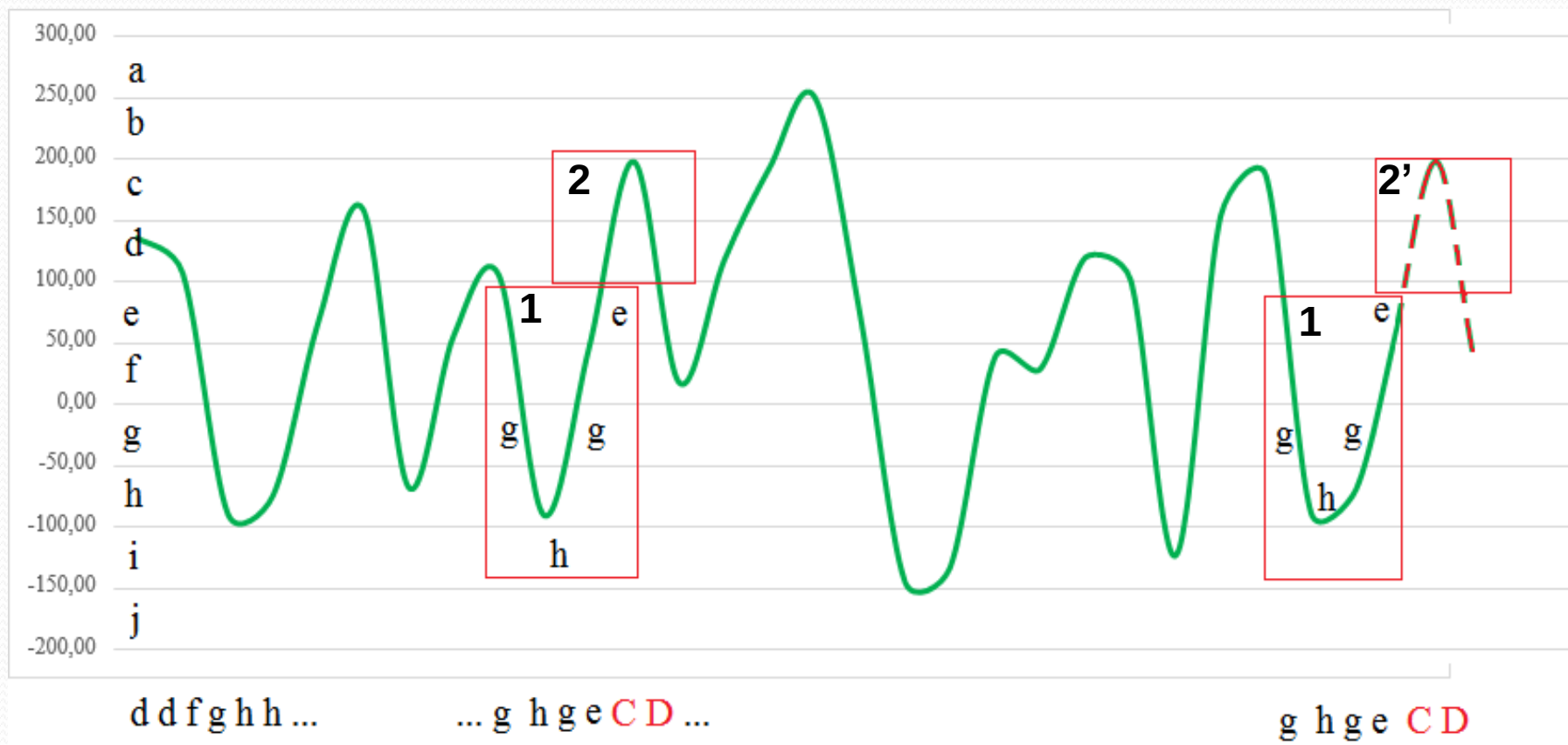
Математичні методи для побудови моделей і прогнозування

Метод подібних траєкторій

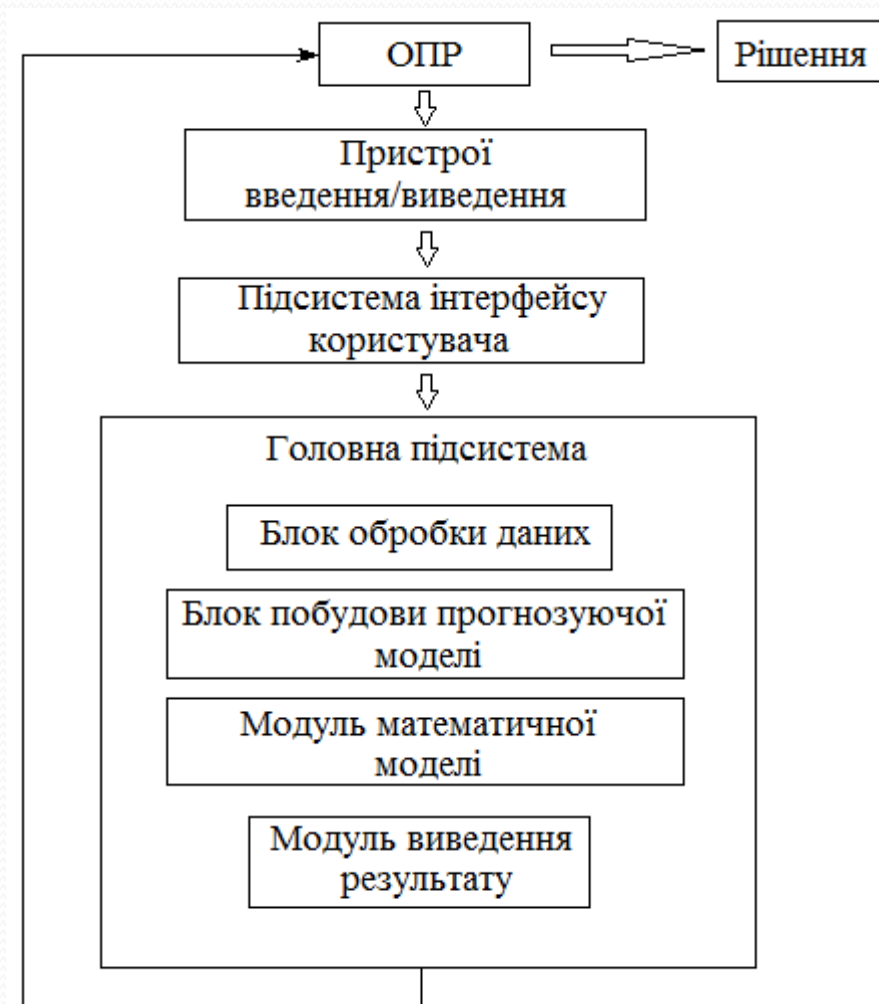


- 1) Обчислимо перші різниці $\Delta y(i) = y(i) - y(i+1), \quad i = \overline{1, N-1}$
- 2) Визначаємо розмір опорного ланцюжка p і розмір алфавіту $B = \{b_k\}$
- 3) Перехід від числового ряду $\Delta y(i)$ до символівного $\langle e_i \rangle$
- 4) Розбиття ряду на вектори розміром p і пошук подібних траєкторій
- 5) Пошук прогнозу – ймовірність появи символу e_{i+1} за умови, що попередніми символами є $e_i, e_{i-1}, \dots, e_{i-k}$

Метод подібних траєкторій



Структура інформаційно-аналітичної системи



Призначення:

- ☐ аналіз даних;
- ☐ прогнозування;
- ☐ прийняття рішень.

Оцінка параметрів моделі

1) Коефіцієнт детермінації

$$R^2 = \frac{\text{var}(\hat{y})}{\text{var}(y)}$$

2) Квадратний корінь із середньоквадратичної похибки

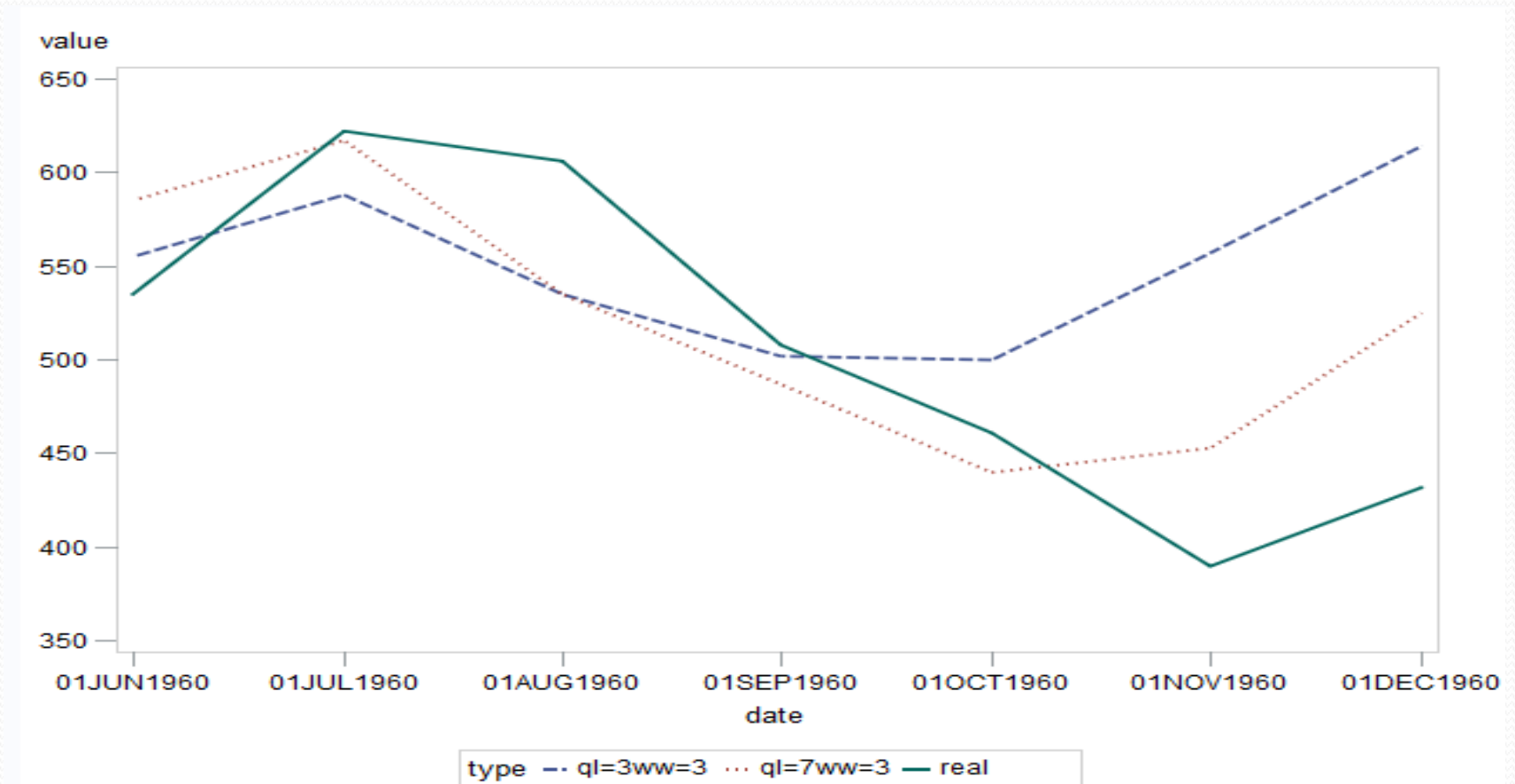
$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}.$$

3) Коефіцієнт Дурбіна-Уотсона

$$DW = \frac{\sum_{t=2}^N (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^N \varepsilon_t^2}$$

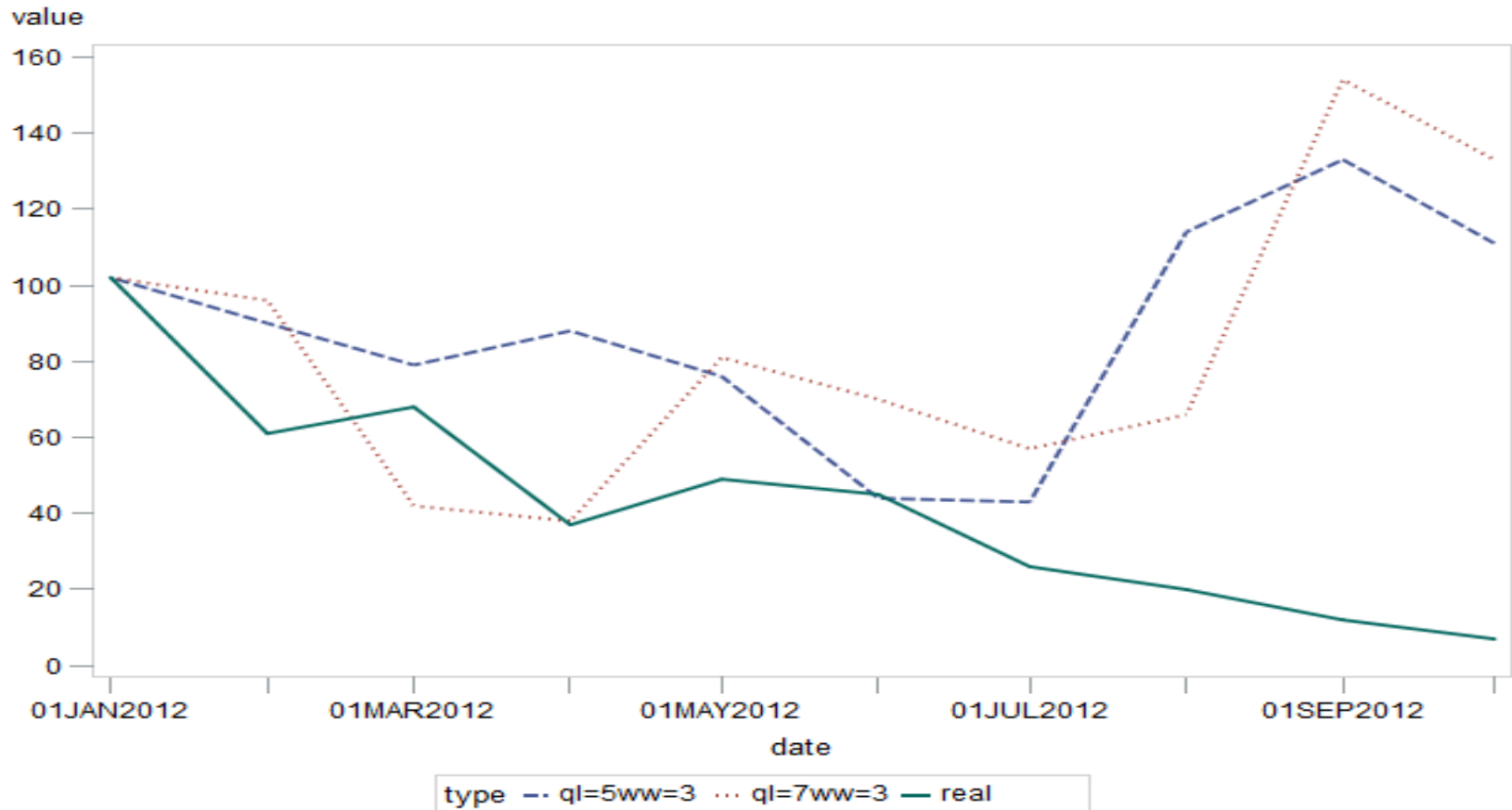
Аналіз результатів

Прогнозування тестового числового ряду SAS AIR



Аналіз результатів

Прогнозування динаміки злочинності у місті Дніпро



Аналіз результатів

Статистичні характеристики побудованих моделей

	RMSE	DW	R^2
Тестовий графік AIR $qI=3$, $ww=3$	9881	0,80	0,34
Тестовий графік AIR $qI=3$, $ww=7$	3009	0,99	0,67
Графік кількості правопорушень у м.Дніпро $qI=5$, $ww=3$	1935	1,03	0,72
Графік кількості правопорушень у м.Дніпро $qI=7$, $ww=3$	960	0,96	0,86

Аналіз результатів

Статистичні характеристики побудованих моделей

	RMSE	DW	R ²
AP(12)	1658	0,97	0,64
АРКС(1,12)	1522	0,93	0,73
Хольт-Вінтерс	1108	0,9	0,81
МПТ(qI=5, ww=3)	1935	1,03	0,72
МПТ(qI=7, ww=3)	960	0,96	0,86

Висновки по роботі

- ❑ Проведений аналіз існуючого інформаційно-аналітичного забезпечення боротьби зі злочинністю;
- ❑ Досліджено ряд математичних моделей для прогнозування соціальних, політичних та економічних процесів;
- ❑ Розроблена архітектура та принципи функціонування інформаційної системи;
- ❑ Здійснено програмну реалізацію частини інформаційної системи на платформі SAS.

Перспективи щодо подальших досліджень

Реалізація нових методів прогнозування в системі
SAS Enterprise Guide:

- ☐ Метод групового врахування аргументів (МГВА)
- ☐ Нечіткий МГВА
- ☐ Когнітивне моделювання

Результати впровадження, публікації

- ❑ Публікація наукової статті “Побудова предиктивної моделі прогнозування злочинності на основі методу подібних траєкторій в системі SAS” (CAIT-2017)
- ❑ Публікація наукової статті “Побудова в системі SAS предиктивної моделі прогнозування злочинності на основі методу подібних траєкторій” (Призер конкурсу наукових робіт SAS 2016)
- ❑ Публікація наукової статті “Аналіз, моделювання та прогнозування злочинності України із використанням методів регресійного аналізу в системі SAS Enterprise Guide” (ISDMCI 18-22 травня 2015 року).
- ❑ Публікація наукової статті “Аналіз злочинності в Україні із використанням методів регресійного аналізу в системі SAS Enterprise Guide” (CAIT-2015)
- ❑ Свідоцтво про реєстрацію авторського права на твір № 71750 від 05.05.17, виданим державною службою інтелектуальної власності України.



Дякую за увагу!