



## ОЦІНЮВАННЯ КЛІОМЕТРИЧНИХ ХАРАКТЕРИСТИК КОНСОЛІДОВАНИХ СТАТИСТИЧНИХ ДАНИХ

Розглянуто нові концептуальні підходи до використання великих масивів даних у соціогуманітарних дослідженнях і проаналізовано сучасні напрями їх використання у науково-дослідній діяльності гуманітарного спрямування. Обґрунтовано актуальність використання технологій Big Data в історичних наукових дослідженнях і використання таких технологій для побудови науково-дослідних кластерів з метою ефективного дослідження й опрацювання масових історичних джерел, а саме зведених статистичних даних. Опрацьовано зведені статистичні дані індустріального розвитку, перепису населення та експертного оцінювання соціо-економічного і демографічного розвитку Сполучених Штатів Америки та Української Радянської Соціалістичної Республіки у 1970-1980-х роках. Апробовано кількісний метод для отримання консолідованої статистики, що дасть змогу виявити особливості статистичної сукупності загалом і в окремих її компонентах, показати регулярність досліджуваних соціально-економічних явищ процесів. Оцінено вагомість консолідованих статистичних даних, опублікованих на вітчизняних та іноземних ресурсах для їх подальшого використання в історичних дослідженнях. Подано у форматі таблиці, змодельовано кліометричні характеристики консолідованих статистичних даних розвитку Сполучених Штатів Америки та України у 1970-х та у 1980-х роках. Обґрунтовано, що кліометричні характеристики консолідованої статистики розвитку Сполучених Штатів Америки та Української Радянської Соціалістичної Республіки у 1970-1980-х роках, які отримують за допомогою моделювання даних економічного розвитку, даних перепису населення, даних національного доходу та зайнятості, експертного оцінювання, можна використати як нове джерело для вивчення актуального історичного періоду.

**Ключові слова:** інформаційні технології; великі дані; консолідована статистика; історія; кліометрія; інтернет; статистичне моделювання; статистика.

### Вступ

Кардинальні зміни у сфері інформаційних технологій зумовлюють експоненціальне зростання кількості цифрової інформації та формують Великі Дані (Big Data). Раціональне застосування Big Data відкриває унікальні можливості для аналітичних досліджень у різних царинах. У період складних соціально-політичних трансформацій особливо актуальним стає використання технологій Big Data в історичних наукових дослідженнях. Побудова науково-дослідних кластерів на підставі технологій Big Data дає змогу проаналізувати історичні дані, зробити ґрунтовні висновки на підставі проведених досліджень, визначити історичні уроки та не допустити помилок у майбутньому.

Нові концептуальні підходи до використання великих масивів даних у соціогуманітарних дослідженнях одним з перших обґрунтував Пітер Друкер. Учений виділив окрему галузь наукової діяльності – цифрові гу-

манітарні науки, що знаходяться на перетині обчислювальних і цифрових технологій та гуманітарних наук. Цифрові гуманітарні науки містять систематичне використання цифрових ресурсів у гуманітарних науках, а також відображення їх застосування [4, 18].

За визначенням К. Лінч, Д. Ленеї, Big Data в інформаційних технологіях – це набір методів і засобів опрацювання структурованих і неструктурованих різномісних динамічних даних великих обсягів з метою їх аналізу та використання для підтримки прийняття рішень [11]. Якщо виходити з такого визначення, то застосування Big Data в історичній науці особливо актуальне, тому що стає унікальним методом і засобом вивчення масових історичних джерел, зокрема зведених статистичних даних [1].

*Об'єкт дослідження* – опрацювання статистичних даних економічного, індустріального розвитку та перепису населення США та України у 1970-1980-х роках.

### Інформація про авторів:

**Химиця Наталія Олексіївна**, канд. істор. наук, доцент, кафедра соціальних комунікацій та інформаційної діяльності.

Email: [nhymytsa@gmail.com](mailto:nhymytsa@gmail.com); <https://orcid.org/0000-0003-4076-3830>

**Устиянович Тарас Остапович**, студент, кафедра соціальних комунікацій та інформаційної діяльності.

Email: [taras.ustyanovych.dk.2017@lpnu.ua](mailto:taras.ustyanovych.dk.2017@lpnu.ua); <https://orcid.org/0000-0002-6323-7924>

**Цитування за ДСТУ:** Химиця Н. О., Устиянович Т. О. Оцінювання кліометричних характеристик консолідованих статистичних даних. Науковий вісник НЛТУ України. 2021, т. 31, № 1. С. 55–61.

**Citation APA:** Khymytsia, N. O., & Ustyanovych, T. O. (2021). Evaluation of consolidated statistical data cliometric characteristics simulated in the r programming language. *Scientific Bulletin of UNFU*, 31(1), 55–61. <https://doi.org/10.36930/40310109>

*Предмет дослідження* – апробація кількісного методу для отримання консолідованої статистики, що дасть змогу виявити особливості статистичної сукупності загалом і в окремих її компонентах, виявити регулярність досліджуваних соціально-економічних явищ та процесів, визначити причинно-наслідкові зв'язки історичних процесів аналізованого набору даних.

*Мета роботи* – оцінити кліометричні характеристики консолідованих статистичних даних соціо-економічного і демографічного розвитку США та Радянської України у 1970-1980-х роках.

Для досягнення зазначеної мети визначено такі *основні завдання дослідження*:

- оцінити вагомість консолідованих статистичних даних, опублікованих на вітчизняних та іноземних ресурсах з використанням мови програмування R (пакети: dplyr, ggplot2, plotly, DT, corrplot) для їх подальшого використання в історичних дослідженнях;
- удосконалити метод опрацювання історичних статистичних даних і підвищити ефективність детермінації причинно-наслідкових зв'язків історичних процесів і явищ.

*Наукова новизна отриманих результатів дослідження* – розроблено новий підхід до опрацювання Великих Даних історичної тематики на підставі автоматизованих засобів мови програмування R; удосконалено використання математичних методів для здобуття знань з великих інформаційних масивів та моделювання причинно-наслідкових зв'язків, кореляцій та тенденцій зростання/зменшення соціо-економічних чинників.

*Практична значущість результатів дослідження* – удосконалено методи аналізу та розроблено характеристики консолідованих статистичних даних завдяки виведенню їх агрегованих кліометричних характеристик. Це особливо важливо у подальших завданнях застосування інформаційно-комунікаційних та обчислювальних засобів для опрацювання великого обсягу кліометричної інформації, адже завдяки мові програмування R та модулям, що дають змогу опрацьовувати Великі Дані, істотно зменшується час, необхідний для повноцінного проведення дослідження.

*Аналіз останніх досліджень та публікацій.* В історичній науці практикують різні підходи до використання Big Data. У США методи та технології Big Data практикував історик Рой Розенцвайг, який був залучений до різноманітних проєктів цифрової історії, зокрема, це вебсайти про історію США, історичне мислення, про історію науки і техніки [2]. У Росії фундаментальні дослідження у цьому напрямі здійснює Л. Бородкін, який подає новий напрям історичної науки – кліодинаміку, що вирішує проблеми моделювання історичних процесів [3]. Кліодинаміка практикує нові підходи до використання математичного моделювання довгострокових соціально-історичних процесів, теоретичної історії, історичної макросоціології, створення та аналізу історичних баз даних, досліджень соціальної еволюції, історичної демографії [11]. Дослідник Пітер Дорн вивчає структуру збирання архіву цифрових досліджень з точки зору метаданих для депонованих наборів даних [17]. М. Польовий використовує математичні методи у дослідженні інтегральних показників соціально-економічного розвитку [13]. С. Голуб у ролі дослідницьких і аналітичних інструментів у кліометричних дослідженнях застосовує моніторингові інформаційні системи [5, 6, 7]. Проблематику Digital history, перспективи застосу-

вання інформаційних технологій та систем опрацювання, збирання контенту історичних вебфорумів вивчають Т. Устиянович та Н. Химиця [8]. Автори розв'язали задачу класифікації жителів Галичини початку – кінця першої половини ХХ ст. за різноманітними характеристиками з використанням методів машинного навчання [19]. В іншому дослідженні розроблено послідовність дій для виявлення і моделювання історіографічної інформації на інтернет-ресурсах та вебфорумах шляхом Web-Scraping, Data mining and Big Data analytics [9].

Однак застосування великих даних у вітчизняній історіографії наразі тільки популяризується, на відмінну від західноєвропейських і американських трендів, які активно використовують такі технології.

*Матеріали та методи дослідження.* Для отримання практичного результату проведеного дослідження застосовано метод синтезу та аналізу джерел інформації, історіографічних даних та метаданих. Використано ресурси мережі Інтернет, банки та бази даних для забезпечення процесу дослідження вхідною інформацією. Пропонуємо автоматизувати збір історичних даних шляхом розроблення інформаційно-пошукового ресурсу з відкритими даними для дослідників історичної науки, який матиме відкритий прикладний програмний інтерфейс або ж API (примітка автора – application programming interface).

У процесі дослідження використано мову програмування R, що дало змогу реалізувати такі методи: візуалізація даних, статистичні обчислення, регресійний аналіз, статистичний аналіз, аналіз співвідношень, автоматизовані методи порівняльного дослідження. Для отримання кліометричних характеристик великих масивів даних використовували інтелектуальний аналіз даних та різноманітні пошукові методи мережі Інтернет, зокрема банк даних World Bank Open Data, з якого було використано всі дані для проведення дослідження [1].

Основним статистичним методом була описова статистика, оскільки опрацьовували саме історичні дані та здійснювали їхнє порівняння з урахуванням різноманітних характеристик. Цей спосіб описової статистики також спрощує знаходження причинно-наслідкових зв'язків та оцінювання кліометричних характеристик. Здійснення кількісного аналізу також необхідне під час опрацювання консолідованих статистичних даних, тому в мові R застосовувався цей метод за допомогою вбудованих функцій. Усі використані модулі мови програмування R є вказані і частково описані у дослідженні. Особливістю мови R є велика кількість вбудованих статистичних та математичних функцій, зручний інтерфейс для опрацювання даних, модулі і пакети для візуалізації та трансформації даних у зведені, виявлення кліометричних характеристик тощо.

## Результати дослідження та їх обговорення

Мова програмування R дає змогу імплементувати лінійну регресію, щоб змодельовати приріст ВВП на душу населення. За допомогою лінійної регресії визначено локальний тренд, а також коефіцієнти зростання, зменшення аналізованих показників упродовж певного історичного періоду.

На рис. 2 змодельовано лінійну регресію приросту ВВП на душу населення у 1970-1980-х роках для США та УРСР. Як ми бачимо, значення  $Y_i$  зростає. Також

спостерігалось зростання чисельності населення (рис. 1). Така тенденція була пов'язана з підвищенням рівня економіки та збільшенням ВВП на душу населення. Проте США розвивалися значно швидше. Це стало основним фактором домінації у "холодній війні", яка тривала до розпаду СРСР.

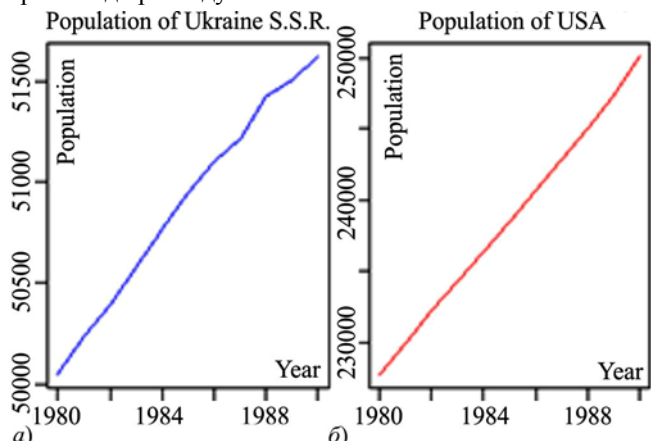


Рис. 1. Населення УРСР та США у 1980-х роках

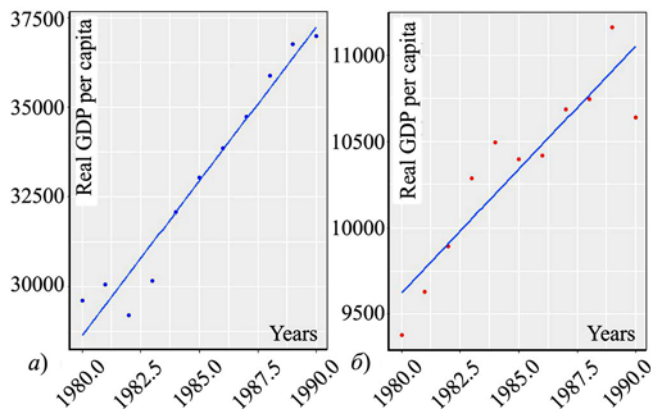


Рис. 2. Реальне ВВП на душу населення США та УРСР у 1980-х роках

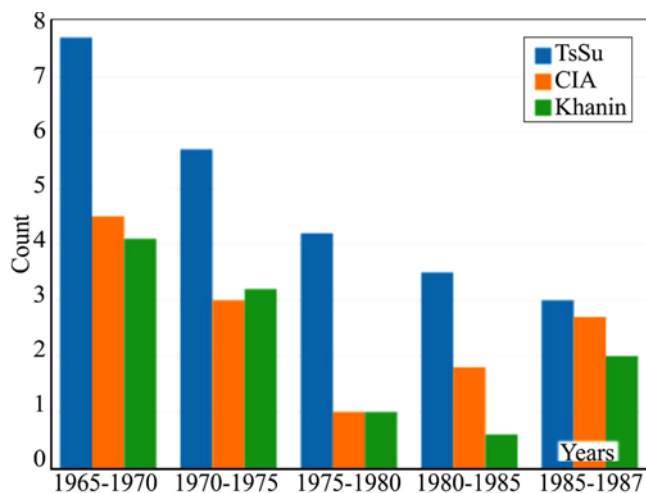


Рис. 3. Дані національного прибутку СРСР упродовж 1965-1987 рр. з різних джерел

Відомо, що радянська статистика є перебільшеною, тому їй не варто довіряти. У графіку на рис. 3 є декілька джерел статистики національного доходу СРСР. ЦСУ (TsSu, Центральне статистичне управління) мало тенденцію до перебільшення рівня економічного розвитку країни. На це вказують дослідження іноземних організацій, які підтверджують наявність неправдивих даних у статистиці СРСР. Російський економіст Георгій

Ханін наводить свою версію статистики прибутку СРСР. ЦРУ (СІА, Центральне розвідувальне управління) проводило безліч досліджень для того, щоб переконалися у вірогідності статистичної інформації СРСР. Дані Ханіна і ЦРУ є схожі.

У дослідженні було знайдено співвідношення даних Національного прибутку СРСР з різних джерел, а саме: ЦСУ, ЦРУ, Ханіна.

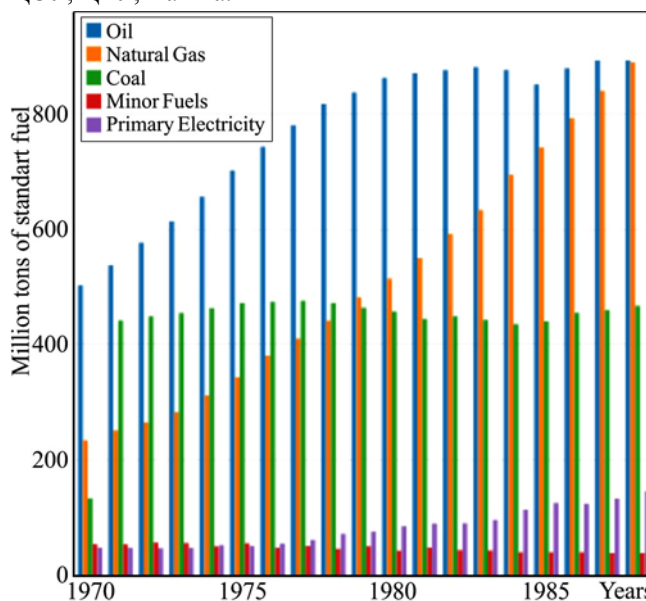


Рис. 4. Загальне виробництво енергії СРСР у період 1970-1987 рр

На рис. 4 показано, що основним прибутком для СРСР у 1970-1980-х роках був експорт нафти. Поступово починає зростати видобуток газу. На початку 1980-х спостерігаємо зниження рівня видобутку вугілля. Порівняно з іншими галузями, менше виробляли електроенергії.

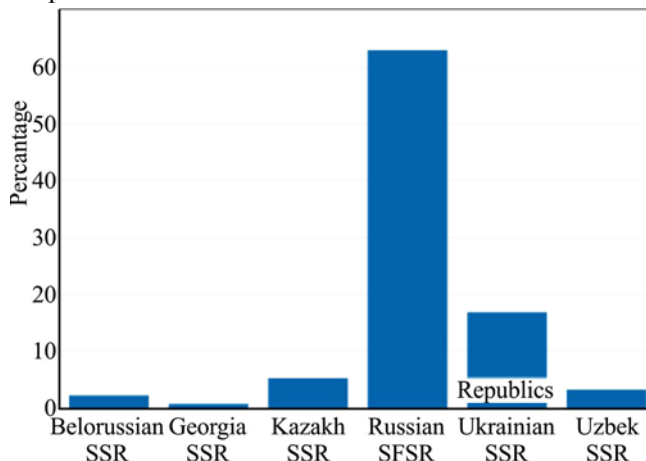


Рис. 5. Виробництво електроенергії по республіках (1987 р.)

УРСР посіла друге місце у виробництві електроенергії в СРСР (рис. 5). За даними 1987 р. вона виробила приблизно 18 % від усіх республік.

У видобутку вугілля УРСР також посіла друге місце (рис. 6). На 1987 р. він становив 25 % від всіх республік. Більшість запасів вугілля зосереджувались на Сході України. У 1991 р. кількість шахт на території України становила 276.

Відповідно до рис. 7, найбільший дохід УРСР на 1972 р. отримувала від промисловості. Це 29, 400 млн рублів. Індустріальний потенціал України збільшився

завдяки побудові великої кількості нових великих заводів (близько 900). Темпи розвитку промисловості УРСР перевищували середні по СРСР. Сільське господарство приносило державі 14, 100 млн рублів. Транспорт та комунікації також відігравали важливу роль в економіці УРСР. Національний дохід від них становив 11,870 млн рублів. Інші галузі також мали значний попит. Їх прибуток становить 22, 270 млн рублів. Найменше прибутків УРСР давало будівництво – 5, 800 млн рублів.

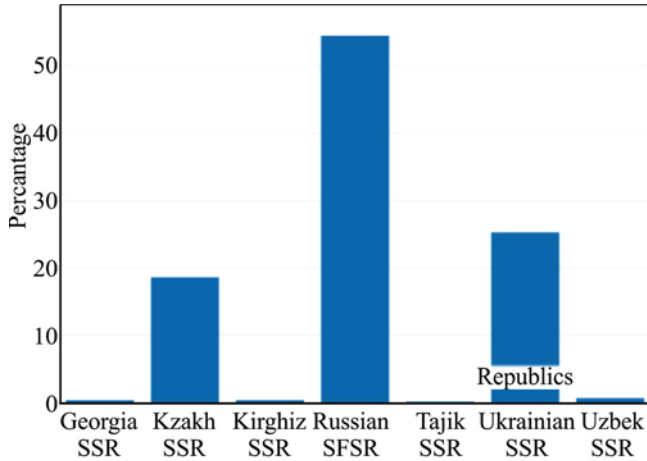


Рис. 6. Видобуток вугілля по республіках (1987 р.)

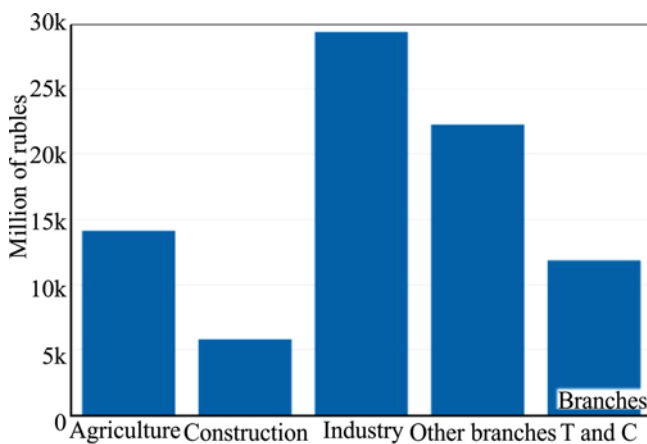


Рис. 7. Національний дохід УРСР (1972 р.)

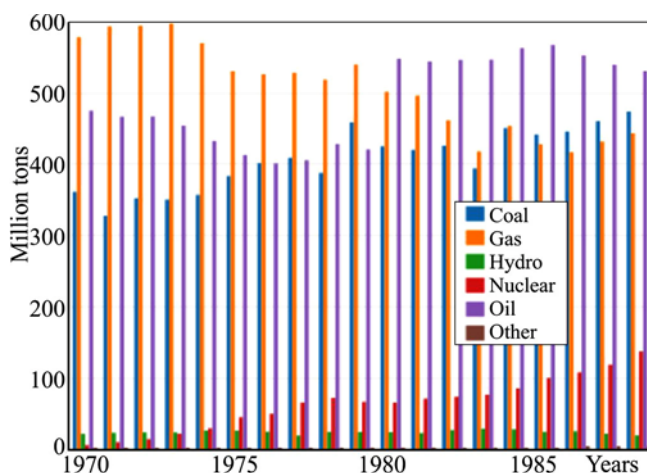


Рис. 8. Загальне виробництво енергії США

У США спостерігалась дещо схожа ситуація. Велике значення в енергетиці мав газ, проте рівень його видобутку почав падати з 1975 р. (рис. 8). У 1980 р. збільшується видобуток нафти. Гідроенергетика та вугільна промисловість мали майже однаковий рівень – 300-400 млн тонн на рік. Виробництво ядерної енергетики

зросло, тоді як виробництво інших видів виробництва електроенергії знизилось.

У XXI ст. нафта і газ стали основними енергоресурсами. Нафтовидобуток у США у 2018 р. досяг рекордного рівня. Відповідні дані оприлюднили в Управлінні енергетичної інформації (ВЕІ) США.

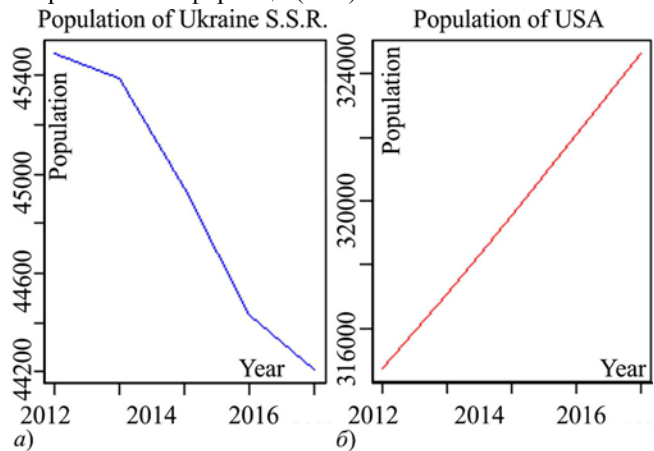


Рис. 9. Населення України та США у другому десятилітті XXI ст.

У другому десятилітті XXI ст. ситуація в Україні залишалася досить нестабільною. Економіка розвивалася повільніше та падала чисельність населення. Це показано на рис. 9.

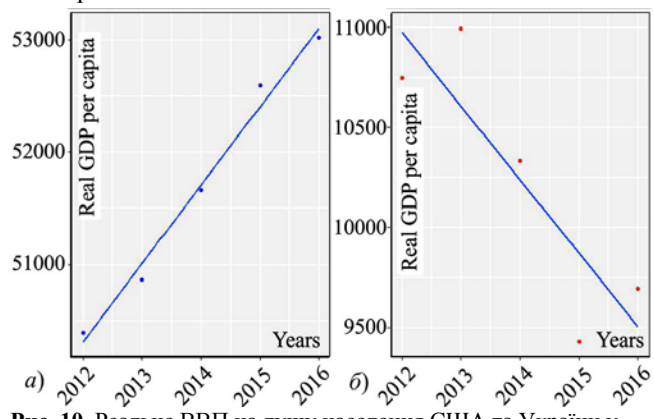


Рис. 10. Реальне ВВП на душу населення США та України у другому десятилітті XXI ст.

Було змодельовано лінійну регресію та отримано значення коефіцієнта нахилу, яке становить для України -3.016, що є низьким показником, тоді як для США – 14.19, що є високим результатом. Окупація Донбасу не дає змогу Україні самостійно розвиватися в енергетичній промисловості, що спричинило зменшення рівня ВВП та неспроможність задовольнити потреби виробництва. В Україні зараз складна демографічна ситуація. Це пов'язано з війною на Сході України, еміграцією частини населення та іншими чинниками. У США відбувається зростання населення, збільшення економічної ваги держави на світовому ринку.

З 2013 р. реальне ВВП на душу населення в Україні починає падати, проте з 2015 р. почалося поступове збільшення. У США воно зростає, що зумовлено швидким розвитком економіки, реформуванням країни, глобалізацією ринку, стабільною монетарною політикою.

Результати опрацювання та моделювання даних економічного та індустріального розвитку, перепису населення та експертного оцінювання соціо-економічного і демографічного розвитку США та УРСР у 1970-1980-х роках подано у таблиці.

**Таблиця. Оцінювання кліометричних характеристик зведених статистичних даних соціально-економічного та індустріального розвитку США та УРСР у 1970-1980-х роках**

Дані, які були опрацьовані та змодельовані (сайти, бази даних)	Оцінювання кліометричних характеристик зведених статистичних даних		
	Пакет, який був використаний для опрацювання та моделювання даних	Результат опрацювання та моделювання даних	Оцінювання кліометричних характеристик
Чисельність населення України та США кінця ХХ – початку ХХІ ст. Реальне ВВП на душу населення у доларах США за 2011 р. (www.data.worldbank.org) (див. рис. 1, 10).	basic plot, plotly, dplyr, ggplot2	Лінійний графік показує позитивну динаміку зростання кількості населення в УРСР та США упродовж 1960-1980 рр. Негативна динаміка зростання населення була в Україні упродовж 2012-2016 рр., тоді як у США – позитивна.	Дана модель дає змогу побачити істотні відмінності між кількістю населення та етапами зростання. Це свідчить про певні причинно-наслідкові зв'язки та потребу їх ретельного вивчення у проаналізовані роки.
Дані Національного доходу СРСР з різних джерел (Радянське економічне зростання з 1928 р. Альтернативна статистика Г. І. Ханіна. Марк Гаррісон) (див. рис. 3)	plotly, dplyr	Національний прибуток СРСР варіюється у різноманітних джерелах, окрім цього, немає чіткого співвідношення і тренду між статистикою, які надають проаналізовані джерела.	Ця модель графіка дає змогу побачити відмінності даних про прибуток СРСР за певні періоди з різних джерел.
Виробництво електроенергії та видобуток вугілля по республіках. Національний дохід Української СР (1972 р.) (див. рис. 6 та 7)	ggplot2	Показано виробництво електроенергії та видобуток вугілля по республіках у відсотках. Згідно з даними, УРСР перебувала на 2 місці в обох категоріях, тоді як видобуток/ виробництво РРФСР був майже вдвічі вищий.	Графіки показуючи загальну ефективність виробництва УРСР для розвитку економіки СРСР.
Загальне виробництво енергії СРСР (див. рис. 4) та США (див. рис. 8).	dplyr, ggplot2, plotly	У СРСР була позитивна динаміка зростання видобування природного газу та нафти; спостерігається щорічне збільшення виробництва первинної електроенергії, тоді як видобування вугілля, мінеральних добрив не має значної динаміки зростання чи спадання. У США є помітне зростання виробництва атомної електроенергії, а також нафти і в окремих періодах вугілля; видобування газу поступово йшло на спад, інші види енергії залишалися незмінні.	Показано основні відмінності між США та СРСР в енергетичній галузі, види палив, які давали змогу економічно розвиватись
Співвідношення даних Національного прибутку СРСР з різних джерел.	corrplot	Дані Ханіна з ЦРУ мають найвище співвідношення (0.92), тоді як знову ж таки Ханіна з ЦСУ – також дещо корелюють (0.85). Тобто між цими наборами є певний схожий тренд. Дані ЦРУ та ЦСУ мають нижчий рівень співвідношення (0.79)	Дає змогу встановити співвідношення даних Національного прибутку СРСР з різних джерел

**Обговорення результатів дослідження.** Ключовим результатом цього дослідження є удосконалення опрацювання історіографічних даних та генерація на їхній основі зведених статистичних даних кліометричних характеристик. У процесі вивчення та аналізу даних було застосовано банк даних World Bank Open Data, який надає у відкритому доступі дані економічного, соціального та індустріального розвитку країн світу у певні історичні періоди.

Для опрацювання даних застосовано мову програмування R, що дало змогу автоматизувати процес очищення та представлення даних, проте використання інших засобів є можливим.

Дослідження кліометричних характеристик дало змогу виявити причину і наслідок боротьби за глобальне лідерство двох економічних гігантів США та СРСР (зокрема, у контексті УРСР) і змоделювати подальший розвиток за певних умов історичного розвитку.

Порівняльна характеристика економічного розвитку СРСР та США дасть змогу виявити особливості історич-

ографічних процесів, встановити причинно-наслідкові зв'язки підсумків Холодної війни (1947-1991 рр.) у певні періоди, а також побачити тенденцію серед різних галузей індустрій. Передусім, варто зазначити, що обидві держави досить активно здійснювали забезпечення енергетичними та паливними ресурсами, проте значення різних джерел даних дещо відрізнялись. Особливістю нашого дослідження є визначення рівня співвідношень між джерелами даних економічного розвитку держав, під час якого було виявлено, що найбільш схожі між собою є дані економіста Ханіна та ЦРУ, менш схожий рівень взаємопов'язаності у ЦСУ.

Схожі порівняльні характеристики було здійснено у праці С. В. Редзюка, де описано економічний світовий розвиток та подано порівняльну характеристику темпів економічного розвитку в Україні та світі [15]. О. Г. Рогожина описує особливості економічної динаміки впродовж 1968-2018 рр. у період активного розвитку технологічних систем масового виробництва, застосовує математичне моделювання довгострокового економічного

розвитку. Здійснено статистичний аналіз показників приросту ВВП та внутрішніх інвестицій в основних регіонах світу (США, Єврозона, Японія) у період 1961(1971)-2017 рр. Розглянуто передумови та наслідки досягнення певних показників у досліджуваних регіонах, результат яких пояснюється гіпотезою про Довгі хвилі Кондратьєва [16].

Не менш важливим та значущим є аналіз стратегії вивчення соціальної історії СРСР в американській історіографії у період з 1980 до початку 2000-х років, здійснений Лаас Н. О., що дає розуміння об'єктивності і можливого рівня упередження США щодо СРСР [10]. Схоже дослідження здійснив Д. Луковицький, у якому йдеться про зміну міжнародних відносин та політики США до України після розпаду СРСР [12].

Пугачова М. В. у своїй статті здійснила порівняльний аналіз розвитку української промисловості ЄС загалом, Польщі, Литви, застосовуючи схожі статистичні методи. Особливістю цієї статті є застосування таких характеристик, як попит, пропозиція, різноманітні коефіцієнти, що покращує ефективність моделювання економічного та індустріального розвитку держав [14]. Для удосконалення статистичного моделювання та аналізу доцільно використати лінійну регресію, що дасть змогу швидко отримати необхідні коефіцієнти й побачити позитивну/негативну динаміку соціально-економічних факторів. Консолідація кліометричних статистичних даних і їх узагальнена характеристика є передумовою для ефективного проведення порівняльного аналізу різних країн у конкретні історичні періоди, що також сприяє виведенню нових знань для історичної науки за допомогою аналізу даних з мережі Інтернет.

## Висновки

Проаналізовано сучасні підходи до використання великих масивів даних у соціогуманітарних дослідженнях. Обґрунтовано актуальність використання технологій Big Data для побудови науково-дослідних кластерів з метою ефективного дослідження й опрацювання масових історичних джерел соціально-економічного та індустріального розвитку. Опрацьовано зведені статистичних дані індустріального розвитку, перепису населення та експертного оцінювання соціо-економічного і демографічного розвитку Сполучених Штатів Америки та Радянської України у 1970-1980-х роках. Апробовано кількісний метод для отримання консолідованої статистики, що дасть змогу виявити особливості статистичної сукупності загалом і в окремих її компонентах, показано регулярність досліджуваних соціально-економічних явищ і процесів. На підставі проведеного дослідження оцінено вагомість консолідованих статистичних даних, опублікованих на вітчизняних і іноземних ресурсах для їх подальшого використання в історичних дослідженнях. Обґрунтовано, що кліометричні характеристики консолідованої статистики розвитку Сполучених Штатів Америки та Радянської України у 1970-1980-х роках, які отримують за допомогою моделювання даних економічного розвитку, даних перепису населення, даних національного доходу та зайнятості, а також експертного оцінювання, можна використати як нове джерело для вивчення актуального історичного періоду.

## References

1. Bank danikh "World Bank Open Data". (2020). Retrieved from: <https://data.worldbank.org/>
2. Bernstein, A. (2007). Digital Historian Roy A. Rosenzweig. *Washington Post*. Retrieved 8, October 2014. Retrieved from: <https://thanksroy.org/items/show/596>.
3. Borodkin, L. I. (2016). Modeling of historical processes: from the reconstruction of reality to the analysis of alternatives *St. St. Petersburg: Aletya*.
4. Drucker, J. (2013). Intro to Digital Humanities: Introduction. *UCL Center for Digital Humanities*. Retrieved December 26, 2016.
5. Golub, S., & Khymytsia, N. (2018, August). The Method of Cliodinamik Monitoring. In *2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)*, 223–226. <https://doi.org/10.1109/DSMP.2018.8478459>
6. Golub, S., & Khymytsia, N. (2016, May). Assessment of utility models in the process of cliometric researches. In *Proceedings of the 5th International Scientific Conference "Information, communication and society"*, 238–240.
7. Holub, S., & Khymytsia, N. (2016, February). The use of multi-level modeling in the cliometric studies process. In *2016 13th International Conference on Modern Problems of Radio Engineering, Telecommunications and Computer Science (TCSET)*, (pp. 733–735). IEEE. <https://doi.org/10.1109/TCSET.2016.7452166>
8. Khymytsia, N., & Ustyianovych, T. (2017, December). Application of big data in historical science. In *VIIIth International Youth Science Forum Litteris et Artibus: proceedings*, (pp. 375–376). Lviv Polytechnic Publishing House.
9. Khymytsia, N., Ustyianovych, T., & Dronyuk, I. (2019). Identification and modeling of historiographic data in the content of web forums. In *CEUR 1st International Workshop on Control, Optimization and Analytical Processing of Social Networks, COAPSN 2019 Proceedings*. Vol. 2392, (pp. 297–308). Lviv, Ukraine.
10. Laas, N. O. (2010). Sotsialna istoriia SRSR v amerykanskkii istoriografii: teoretychni dyskusii 1980-2000-kh yahr. *Ukrainskyi istorychnyi zhurnal*, 5, 170–191. Retrieved from: [http://resource.history.org.ua/publ/journal\\_2010\\_4\\_170](http://resource.history.org.ua/publ/journal_2010_4_170). [In Ukrainian].
11. Laney, D., & Mark, A. (2012). The Importance of "Big Data": A Definition. *Gartner Research*. Retrieved from <https://www.gartner.com/doc/2057415/importance-big-data-definition>.
12. Lukovytskyi, D. (2018). Stavlennia SShA do ruinatsii SRSR: zmina vidnosyn z Ukrainoiu (1986–1994 rr.). [In Ukrainian].
13. Polyvovyy, N. (2014). Simulation Modeling of Political Instability and Maydan of 2013/2014 in Ukraine. *European Journal of Transformation Studies*, 2, 22–33.
14. Puhachova, M. V. (2017). Statystychnyi analiz rozvytku ukrainskoi promyslovosti: chy ye rezervy? *Statystyka Ukrainy*, 3, 20–27. Retrieved from: [http://nbuv.gov.ua/UJRN/su\\_2017\\_3\\_5](http://nbuv.gov.ua/UJRN/su_2017_3_5). [In Ukrainian].
15. Redziuk, Ye. V. (2013). Porivnialna kharakterystyka tempiv ekonomichnoho zrostantia ta kapitalizatsii fondovykh birzh svitu y Ukrainy. *Finansy Ukrainy*, 7, 78–90. [In Ukrainian].
16. Rohozhyn, O. H. (2018). Osoblyvosti dovhostrokovoï ekonomichnoi dynamiky za ostanni 50 rokiv. *Matematychni modelivannia v ekonomitsi*, 5, 50–63. [In Ukrainian].
17. Schamhorst, A., Bosch, O. T., & Doom, P. (2012). Looking at a digital research data archive-Visual interfaces to EASY. *arXiv preprint arXiv:1204.3200*.
18. Terras, M. (2011). Quantifying Digital Humanities. *UCL Centre for Digital Humanities*. Retrieved December 26, 2016.
19. Ustyianovych, T., & Khymytsia, N. (2018). Application of machine learning algorithms for the study of Galicia populations social characteristics in the interwar period (1919-1939). In *VIIIth International Youth Science Forum Litteris et Artibus: proceedings*, (pp. 96–101). Lviv Polytechnic Publishing House.

## **EVALUATION OF CONSOLIDATED STATISTICAL DATA CLIMETRIC CHARACTERISTICS SIMULATED IN THE R PROGRAMMING LANGUAGE**

New conceptual approaches to Big Data usage in socio-humanitarian research are considered and their current usage directions in the humanities research activities are analyzed. The relevance of Big Data technologies in historical research is substantiated, in particular as an efficient historical data source studying method, namely consolidated statistical data. Using the R programming language (packages: dplyr, ggplot2, plotly, DT, corrplot), the data and summary statistics of industrial development, census, and expert assessment of socio-economic and demographic development of the United States of America and the Ukrainian Soviet Socialist Republic in the 1970s – 1980s were processed. A quantitative method for obtaining consolidated statistics has been tested, which will reveal statistical population features as a whole and in its individual components, show the regularity of the studied processes socio-economic phenomena. On the basis of the conducted research, the consolidated statistical data weight estimation published on domestic and foreign resources for their further use in historical researches is made. The cliometric characteristics of the consolidated statistical data of the United States of America and the Ukrainian Soviet Socialist Republic in the 1970s – 1980s development, modeled in the R programming language, are presented in the visual table and graphic formats. It is substantiated that the cliometric characteristics of the consolidated statistics of the United States of America and the Ukrainian Soviet Socialist Republic in the 1970s – 1980s development, which are obtained by statistical modeling can be used as a new source for the current historical period studies.

**Keywords:** information technology; Big Data; consolidated statistics; history; cliometrics; the Internet; statistical modeling; statistics.