

РОЛЬ ЗАСОБІВ МАСОВОЇ ІНФОРМАЦІЇ В ПРОЦЕСІ ФОРМУВАННЯ ІНФЛЯЦІЙНИХ ОЧІКУВАНЬ

ТЕТЯНА ЮХИМЕНКО^a

^aНаціональний банк України

Електронна пошта: tetiana.yukhymenko@bank.gov.ua

Анотація

У дослідженні розглядається роль ЗМІ у формуванні інфляційних очікувань серед різних респондентів в Україні. На основі великої сукупності новин із використанням методів машинного навчання отримано відповідні виміри. Вони подають кількісні показники для текстів, які визначають, чи теми новин відповідають інфляційним очікуванням. Дослідженням виявлено докази того, що різні теми новин можуть впливати на інфляційні очікування та здатні пояснити частину їх розбіжності. Результати дослідження можуть допомогти під час аналізу інфляційних очікувань, що є цінним, ураховуючи, що закріплення інфляційних очікувань залишається ключовим завданням для центральних банків.

Класифікація JEL

C55, C82, D84, E31, E58

Ключові слова

інфляційні очікування, обробка природної мови, текстові дані, машинне навчання

1. АКТУАЛЬНІСТЬ, ТЕОРЕТИЧНА БАЗА, ОГЛЯД ЛІТЕРАТУРИ

Закріплення інфляційних очікувань залишається ключовим викликом для центральних банків, особливо в країнах, що розвиваються. Процес формування інфляційних очікувань важливий для розуміння макроекономічної динаміки та розроблення оптимальних варіантів економічної політики. Проведено чимало досліджень у сфері інфляційних очікувань, але все ще є багато невизначеності та суперечливості щодо чинників, які їх визначають. Розвиток сучасних інформаційних технологій дає можливість використовувати нові підходи до вивчення процесів, що формують інфляційні очікування. Зокрема, засоби оброблення природної мови та машинного навчання можуть надати додаткову інформацію, яка раніше була недоступною. Вони також дають змогу доповнити новими ідеями результати наявних досліджень, які стали еталоном у галузі. Зараз багато дослідників звертається до сучасніших джерел даних і методів аналізу, але галузь дослідження залишається недостатньо вивченою. Зокрема, залишається велика невизначеність щодо того, як трансформувати неструктуровані дані в економічні показники, як урахувувати тональність показників і як оцінювати

їхній вплив на інфляційні очікування. Крім того, на базі українських даних такі дослідження ще не проводилися. Таким чином, перспектива застосування новітніх технологій до вже традиційних підходів була основною мотивацією для дослідження ролі ЗМІ у формуванні інфляційних очікувань в Україні.

Гіпотеза раціональних очікувань домінувала в макроекономічній літературі протягом багатьох років. Однак у дедалі більшому обсязі досліджень ця гіпотеза модифікується для врахування інформаційної жорсткості – очікування можуть бути раціональними, але в більш реалістичному середовищі економічні учасники можуть бути неуважними до відповідної інформації через витрати на отримання та оброблення такої інформації. Двома провідними моделями інформаційної жорсткості є повільно змінювана інформаційна модель Mankiw et al. (2004) та шумова інформаційна модель, розроблена Woodford (2004) і Sims (2009). Mackowiak and Wiederholt (2009) також зробили свій внесок у цій сфері. Coibion and Gorodnichenko (2012) довели, що інформаційна жорсткість має великий вплив на макроекономічні змінні, тому їх слід інтегрувати в сучасну макроекономічну політику, щоб здійснювати оптимальну монетарну політику. Вони також виявили, що, незважаючи на загальну думку, немає істотної різниці в ступені споживання інформації між учасниками економічних відносин – швидкість

оброблення інформації споживачами не нижча, ніж в інших учасників. Зокрема, це можна пояснити шумовою інформаційною моделлю. Подібним чином Coibion and Gorodnichenko (2015a) виявили, що інфляційні очікування професійних прогнозистів з U.S. Survey можна моделювати за допомогою моделей неповної інформації через існування так званого інформаційного тертя. Coibion and Gorodnichenko (2015b) також досліджують очікування економічних суб'єктів в Україні на основі даних опитувань щодо інфляції та очікувань обмінного курсу. Опитування також показує, що існує сильна позитивна кореляція між еволюцією очікувань українських економічних агентів (учасників економічних відносин) щодо інфляції та курсу валют. Хоча можна очікувати певної кореляції від переносу обмінних курсів на ціни, більш вірогідним обґрунтуванням є те, що курс валют використовується домашніми господарствами як прямий індикатор ширших коливань цін в економіці, і це дуже схоже на те, як в домашніх господарствах США змінюються ціни на паливо.

Можна припустити, що на респондентів опитування також впливає невизначеність щодо податкової, тарифної, бюджетної, монетарної та регуляторної політики. Ці ефекти, однак, важко виявити, оскільки невизначеність не спостерігається. Проте люди можуть прямо чи опосередковано формувати свою думку про майбутнє економіки зі ЗМІ. Таким чином, методи на основі новин можна використати для дослідження впливу медіасередовища на формування очікувань респондентів.

Наприклад, Carroll (2003) провів тестування епідеміологічної моделі очікувань, у якій інформація з часом поширюється від професійних прогнозистів до домогосподарств. Pfajfar and Santoro (2013) доповнюють цю модель вимірюванням фактичного сприйняття нової інформації про ціни. Як кількісні показники новин вони використовували запитання з опитування, де учасники мали зазначити, чи чули вони про позитивні чи негативні зміни. Отримання новин, пов'язаних із цінами, підвищує можливість коригування інфляційних очікувань, тоді як якість прогнозів навряд чи поліпшиться. Подібним чином Coibion et al. (2019) досліджували, як повідомлення центрального банку впливають на очікування. Таким чином, вони порівнюють відповіді респондентів після отримання восьми різних форм інформації щодо інфляції. Вони дійшли висновку, що ці повідомлення громадськості впливають на очікування в економічно значних масштабах. Однак їх ефективність значно знижується під час передавання через засоби масової інформації. Mazumder (2021) довів, що газетні згадки про ФРС зближують споживчі та професійні прогнози інфляції, хоча цей ефект може відрізнитися залежно від того, в якій газеті вони були опубліковані та як саме тема висвітлювалася автором. Dräger and Lamla (2017) також знайшли докази впливу ЗМІ на формування інфляційних очікувань. Вони проаналізували мікродані "Опитуванні споживачів" Мічиганського університету та знайшли докази того, що респонденти з більшою ймовірністю коригуватимуть свої очікування, якщо почують новини про інфляцію.

Однак більшість цих досліджень передбачає використання додаткових запитань в опитуванні, що може зробити всі ці опитування надто дорогими в сенсі коштів. Крім того, навіть якщо такі запитання

будуть запроваджені, то результати не охоплюватимуть попередніх періодів. Таким чином, вимірювання впливу новин і розрахунок відповідних показників потребує нових джерел інформації та методів оброблення, а також значних обчислювальних ресурсів. Отже, дослідники заміняють ці показники альтернативними показниками, які можуть бути пов'язані з кількісними показниками новин. Наприклад, Bauer (2015) використовував неочікувані відхилення макроекономічних даних, накопичені протягом місячних або кварталних періодів спостереження, як показник економічних новин. Таким чином, дані є макроекономічними показниками, зібраними з традиційних статистичних джерел, але їх інтерпретація дещо відрізняється від звичайних часових рядів. Бауер виявив, що кілька різних вимірювань інфляційних очікувань значною мірою реагують на макроекономічні неочікувані відхилення. Він також дійшов висновку, що краще закріплення довгострокових інфляційних очікувань може зменшити чутливість інфляційних очікувань до макроекономічних новин, а також до мінливості номінальних ставок. Garcia and Werner (2018) підтвердили, що раннє оприлюднення інфляційних даних мало значний вплив на довгострокові інфляційні очікування і що в останні роки відбулося послаблення закріплення інфляційних очікувань в ЄС. Nautz et al (2017) також виявили, що закріплення інфляційних очікувань єврозони було порушене після осені 2011 року. Вони виявили, що довгострокові інфляційні очікування значною мірою реагують на макроекономічні новини. Як кількісні показники новин використовувався набір макроекономічних змінних, включаючи індекс споживчих цін, індекс цін виробників, безробіття, ВВП, торговий баланс тощо. D'Acunto et al. (2017) додатково виявили залежність між частотою та розміром зміни цін.

Larsen et al. (2021) використали більш складний підхід. Вони застосували алгоритми машинного навчання до великого корпусу новин і дослідили роль ЗМІ в процесі формування очікувань домогосподарств. Виявилось, що тематика новин у ЗМІ є хорошим провісником як інфляції, так й інфляційних очікувань. Вони також виявили, що ступінь інформаційної жорсткості серед домогосподарств змінюється з часом, що можна пояснити відповідним висвітленням у ЗМІ. Angelico et al. (2021) використали подібний підхід для побудови вимірювань інфляційних очікувань споживачів у реальному часі на основі записів у соціальній мережі мікроблогів "Twitter". Вони поєднали неконтрольовані методи машинного навчання з підходом на основі словника для розрахунку показників. Індикатори на основі Twitter, як видається, сильно корелюють із традиційними вимірюваннями інфляційних очікувань, але мають перевагу в швидкості.

У цій праці ми приділяємо увагу аналізу новин та їх впливу на формування інфляційних очікувань. З цією метою ми досліджуємо підходи до перетворення текстів на кількісні показники, які потім можна використовувати в традиційному економічному аналізі. Ці показники мають відображати теми новин, пов'язані з інфляційними очікуваннями, і точно відображати їхню інтенсивність. Ці кількісні показники також мають легко інтерпретуватися, оскільки вони спрямовані на пояснення впливу новин на формування інфляційних очікувань. Усі ці завдання можна вирішити за допомогою методів аналізу тексту.

Усі методи вимірювання, які ґрунтуються на текстовому аналізі, можна розділити на дві групи: 1) так звані наївні методи та 2) більш складні методи, засновані на машинному навчанні. Наївні методи прості, легкі у використанні (оскільки не потребують великої обчислювальної потужності) і визнані в усьому світі завдяки своїй простоті. Вони мають за основу переважно частоту термінів і частоту документів. Наприклад, Baker et al. (2016) досліджували зв'язок між невизначеністю економічної політики та темпами зростання інвестицій, виробництва та зайнятості. Для цього автори розробили показник невизначеності економічної політики (НЕП) на основі щомісячного підрахунку статей, які містять специфічні терміни. Результати їхнього дослідження показують, що НЕП є прийнятним показником для різних типів важливих макроекономічних змінних, і його результати узгоджуються з теоріями, які підкреслюють негативні економічні наслідки шоків (потрясіння) невизначеності.

Однак ці підходи потенційно можуть недооцінити фактичний рівень невизначеності, оскільки вони потребують якісної експертизи та людських ресурсів. Наприклад, більшість наївних методів передбачає створення словника. Цю проблему можна вирішити більш складними методами, заснованими на техніці машинного навчання. Незважаючи на відносну складність, підходи машинного навчання мають більшу прогностичну силу, ніж наївні методи, показують емпіричні результати. Найшвидший і найпростіший метод – використовувати техніку машинного навчання без учителя.

Одним із найпопулярніших неконтрольованих інструментів оброблення природної мови є латентне розміщення Діріхле (ЛРД), представлене Blei et al. (2003). Ця генеративна статистична модель поділяє колекцію текстів на підгрупи, причому кожна підгрупа характеризується ключовими словами, пов'язаними з темою. Цей метод оцінює вірогідність появи слів для різної кількості тем. Результати показують найбільш ймовірну кількість тем. ЛРД – це техніка машинного навчання без учителя, яка не потребує набору даних для навчання. Однак результати моделі непередбачувані й потребують ретельного аналізу. Варто зазначити, що в той час як методологія широко застосовувалася в літературі з машинного навчання та аналізу тексту, в економіці, на диво, вона досі мала лише невелику кількість успішних застосувань, наприклад, Larsen et al. (2021) або Azqueta-Gavaldon (2017). Tobbäck et al. (2016) обрали інший шлях і спробували поліпшити перший показник НЕП, розроблений Бейкером, за допомогою застосування керованого машинного навчання. Таким чином, вони розробили модель класифікації на основі методу опорних векторів (МОВ) і розділили статті на два класи – пов'язані чи ні з невизначеністю економічної політики. Крім того, вони визначили показник НЕП МОВ на основі цієї класифікації та включили його в різні макроекономічні моделі. Це допомагає підвищити точність прогнозів економічних змінних цих моделей у короткостроковій перспективі.

Однак неконтрольовані моделі, як і наївні методи, також мають недоліки – немає аналізу емоційного забарвлення висловлювань. Це можна пом'якшити за допомогою методів машинного навчання, що використовують лексикон – заздалегідь визначений словниковий запас, та оцінюють відносну частоту емоційно забарвлених слів у тексті. Наприклад, Taboada

et al. (2011) представили калькулятор семантичної орієнтації (SO-CAL). У цій моделі використовуються словники слів, що характеризуються їх семантичною полярністю та силою з інтенсифікацією та запереченням. SO-CAL можна використовувати для даних, які не зустрічалися раніше. VADER (словник з урахуванням валентності для міркувань про емоційно забарвлені повідомлення) – це один успішний приклад інструменту аналізу емоційно забарвлених повідомлень на основі лексики. Щоб розробити його, Hutto and Gilbert (2014) склали список лексичних ознак та об'єднали їх із загальними вимогами, що втілюють граматичні та синтаксичні правила для вираження інтенсивності емоційно забарвлених повідомлень. VADER перевершив багато інших високо оцінених інструментів аналізу емоційно забарвлених повідомлень. Однак основний недолік методів, заснованих на лексиконі, полягає в тому, що немає підготовлених словників іншими мовами, крім англійської.

Запровадження нової моделі представлення мови під назвою BERT (двоспрямовані кодувальні представлення з трансформерів), розробленої дослідниками корпорації “Google” (Devlin et al., 2018), стало значним проривом в аналізі емоційно забарвлених повідомлень. Як і багато інших нещодавніх робіт із попереднього навчання контекстних представлень, BERT використовує механізм уваги, який вивчає контекстні зв'язки між словами (або частинами слів) у тексті. Але на відміну від багатьох інших моделей, вона розроблена для попереднього навчання глибоких двоспрямованих репрезентацій тексту без міток (оброблення як лівого, так і правого контексту). У результаті цього BERT може розрізнити відмінності у вживанні навіть одного і того самого слова з урахуванням контексту його появи. Попередньо підготовлену модель BERT можна налаштувати для широкого кола завдань, включаючи класифікацію. Попередньо підготовлені версії BERT доступні багатьма мовами (включаючи українську та російську).

Щоб звузати межі цієї статті, ми зосереджуємося на простіших наївних методах і машинному навчанні без учителя, а аналіз емоційно забарвлених повідомлень залишаємо для майбутніх досліджень. Починаючи з найпростіших наївних методів, ми продовжимо з більш складними методами машинного навчання класифікації тексту, такими як ЛРД. Відповідно ми розробляємо економетричну модель для оцінки впливу визначених показників на формування інфляційних очікувань.

Стаття структурована таким чином: у наступному розділі представлені характеристики даних, розділені на дві частини – текстове зведення економічних новин та інфляційних очікувань в Україні. Розділ 3 описує визначення та результати застосування показників на основі новин і представляє їхні статистичні властивості. Розділ 4 аналізує емпіричні характеристики моделей та описує результати. Нарешті в розділі 5 пропонуються деякі заключні зауваження та майбутні кроки. Додаткову інформацію та результати можна знайти в додатках.

2. ХАРАКТЕРИСТИКИ ДАНИХ

2.1. Сукупність новин

Загальними ключовими критеріями для відбору джерел новин були наявність досить великого архіву (принаймні за останні десять років) і можливість

скопіювати дані з мережі, що суттєво обмежувало доступний список. Крім того, використовувані газети повинні мати переважно економічну спрямованість і не піддаватися явному впливу окремих політичних сил. Ми керувалися списком найпопулярніших ресурсів, з якого виключили ті, що не відповідали зазначеним вимогам. Зокрема, ми обрали такі джерела, як “Українська правда”, “Ліга” та УНІАН, котрі входять до списку “ТОП-50 українських інтернет-ЗМІ” за рейтингами Gemius та TNS (6, 11 та 20 місця відповідно). Крім того, ми включили дані сайту “Finance.ua”, який вузько спеціалізується на економічній тематиці, а також увійшов до рейтингу двохсот найпопулярніших сайтів України (за версією alexa.com). За даними Similarweb.com, ці сайти загалом охопили близько 8.5% трафіку онлайн-новин та ЗМІ в Україні в серпні – жовтні 2022 року. Однак ці новинні сайти не надають інформації про кількість переглядів або висвітлення кожної новини. Ця інформація може бути корисною для визначення сили впливу новин у формі множника.

Державною мовою в Україні є українська, але російська мова була дуже популярна до повномасштабного вторгнення росії, тому більшість національних ЗМІ публікували матеріали двома мовами. Водночас інфраструктура оброблення природної мови для російської мови була трохи краще розвинена на момент проведення дослідження, містила багатші бібліотеки та модулі, які могли б поліпшити результати дослідження. З цієї причини було вирішено опрацювати новинні статті російською мовою.

Сукупність новин, зібрана в мережі, складається з понад 2 мільйонів статей, опублікованих онлайн за 20 років – із січня 2000 року до грудня 2020 року. Наш набір даних містить повний текст статті та доступні метадані, які включають, наприклад, дату, посилання та назву. Деякі джерела також мають підзаголовки або загальну тему. Уся робота з текстовими даними, починаючи від вебскрепінгу та попереднього оброблення і закінчуючи визначенням показника, виконувалася мовою програмування Python.

Текстові дані представлені в нетрадиційному форматі, що ускладнює статистичні висновки. Таким чином, важливо попередньо обробити дані сукупності новин в машиночитаний формат. Попереднє оброблення передбачає деякі кроки для очищення та зменшення необробленого набору даних перед аналізом.

По-перше, важливо писати всі символи в нижньому регістрі, щоб уникнути оброблення з урахуванням регістру. Це повинно допомогти очистити набір даних принаймні у двох випадках:

- слова з великими літерами можуть не розпізнаватися як стоп-слово, оскільки всі списки стоп-слів зазначені малими літерами. Стоп-слова – це слова, які не мають істотного внеску в зміст тексту. Наприклад, найпоширенішими стоп-словами є сполучники та прийменники;
- з граматичних міркувань одне і те саме слово може трактуватися по-різному через його позицію в реченні. Наприклад, перше слово в реченні завжди пишеться з великої літери, навіть якщо воно не є власною назвою. У результаті одному слову можна надати два різних значення.

Другим, але фундаментальним кроком у методах НЛП є токенизація (лексемізація). Токенизація – це метод розбиття частини необробленого тексту на менші одиниці, які називаються лексичними символами, і перетворення їх у список. Лексичні символи можуть бути словами, символами або частинами слів. Лексичний символ – це одиниця, яку інструменти НЛП можуть легко перетворити на значення, придатне для подальшого машинного оброблення.

По-третє, ми видалили із тексту несуттєву інформацію, як-от стоп-слова, щоб спростити оброблення даних. Бібліотека NLTK (набір інструментів природної мови) в Python має широкий перелік стоп-слів різними мовами, включаючи російську. Крім того, ми видалили символи, які не належать до ASCII (Американський стандартний код для інформаційного обміну), посилання та знаки пунктуації за допомогою регулярних виразів (Regex).

Четвертий крок попереднього оброблення тексту – нормалізація тексту. Найпоширенішими методами нормалізації для оброблення природної мови є коріння і лематизація. Стемінг – це техніка, яка обрізає кінці слів. Завдяки такому підходу нормалізуються слова з однаковим значенням, але які мають деякі варіації відповідно до контексту чи речення. Лематизація зазвичай належить до морфологічного аналізу слів і здебільшого спрямована на повернення базової або словникової форми слова (леми). Російська мова, як і українська, є морфологічно багатою мовою, яка характеризується вільним порядком слів і різноманітністю словоформ. Майже всі частини мови позначаються багатьма характеристиками, такими як число, рід, відмінок, час, вид чи особа, які граматично узгоджуються між собою (Rozovskaya and Roth). Таким чином, незважаючи на довший час оброблення та потребу в більшій обчислювальній потужності, лематизація є кращим методом, ніж стемінг, для нормалізації російськомовного тексту. Для цього ми використали морфологічний аналізатор rymorphu2, який повертає словникову форму слова (Korobov, 2015).

Врешті-решт ми отримали очищений і нормалізований набір текстових даних, що складається приблизно з 300 мільйонів слів і майже 800,000 унікальних лексичних символів. Як показано на рисунку 1, середній розмір статті не змінився суттєво протягом періоду спостереження (100–140 слів на статтю). Однак кількість статей значно зросла – з кількох статей на місяць на початку 2000-х років до 8,000–14,000 на місяць із 2008 року, а розподіл кількості статей між онлайн-джерелами також змінився (додаткову інформацію про сукупність новин див. у додатку А). Збільшення кількості онлайн-статей у середині 2000-х пов’язане зі швидким зростанням проникнення інтернету в Україні. Так, за даними Державної служби статистики України, у 2007 році кількість активних користувачів інтернету в Україні вперше перевищила 1 млн осіб, поступово збільшуючись із 200 тис. у 2000 році. З того часу кількість активних користувачів зросла до понад 23 мільйонів, і, відповідно, проникнення інтернету зросло до 56% у 2019 році. Це змусило ЗМІ перейти від традиційних паперових форм до онлайн-версій. У результаті ЗМІ не тільки повністю перевели свої статті в інтернет, але й розширили контент сайтів додатковими матеріалами, які зазвичай не розміщуються в газетах.

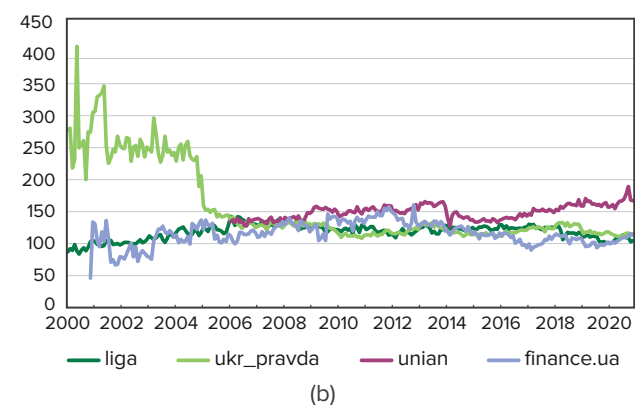
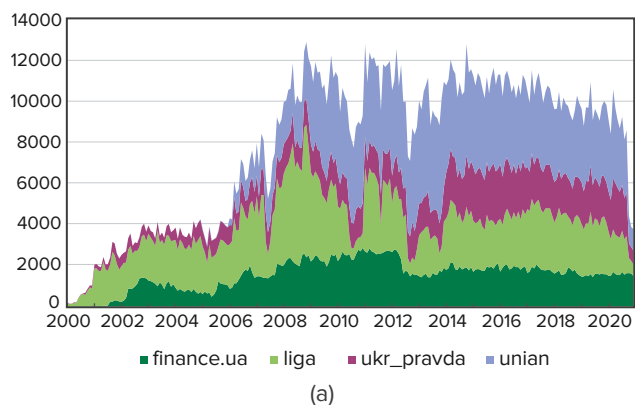


Рисунок 1. Кількість новин (а) і середній розмір статей (б) на місяць

2.2. Інфляційні очікування

Національний банк України проводить дослідження щодо інфляційних очікувань на наступні 12 місяців для кількох типів учасників економічних відносин: домогосподарств, банків, суб'єктів підприємницької діяльності та професійних прогнозистів. До запровадження НБУ режиму встановлення цільових показників інфляції (інфляційне таргетування) у 2015 році Coibion and Gorodnichenko (2015b) детально проаналізували ці опитування та обговорили їхні обмеження. У своїй праці ми коротко опишемо характеристики інфляційних очікувань усіх груп респондентів (рисунок 2).



Рисунок 2. Інфляційні очікування на наступні 12 місяців, %

Джерело: НБУ, GfK Ukraine, Info Sapiens.

Банки. Огляд банків охоплює не менше 90% активів банківської системи, за винятком неплатоспроможних

банків та банків, що перебувають у процесі ліквідації. НБУ почав опитування банків у 2012 році, дані доступні поквартально. Банки опитуються протягом перших тижнів кварталу.

Суб'єкти підприємницької діяльності. Таке опитування містить відповіді близько 700 підприємств нефінансового сектору. Підприємства відбираються за квотним принципом відповідно до структури економіки України, що забезпечує репрезентативність вибірки. Опитування суб'єктів підприємницької діяльності проводяться НБУ щокварталу з лютого 2006 року, однак вони проводяться протягом другого місяця кварталу.

Фінансові аналітики. З липня 2014 року НБУ розпочав щомісячне опитування професійних прогнозистів (протягом другого та третього тижнів кожного місяця). З листопада 2019 року частоту цього опитування скоротили до восьми разів на рік, щоб відповідати графіку засідань Комітету з монетарної політики. Відповіді фінансових аналітиків збираються за тиждень до дати зустрічі. Кількість професійних прогнозистів змінюється з часом – від шести до 12.

Домогосподарства. Одночасно з опитуваннями фінансових аналітиків у липні 2014 року започатковано опитування домогосподарств. На відміну від інших опитувань, опитування домогосподарств щомісяця проводить стороння компанія Info Sapiens (до 2019 року проводила GfK Ukraine). Кожного другого та третього тижня місяця приблизно 1,000 споживачів дають відповіді щодо їхніх інфляційних очікувань та багатьох інших соціальних та економічних питань. Вибірка є національно репрезентативною та змінюється щомісяця.

Банки, підприємства та домогосподарства обирають інтервал очікуваної інфляції на наступні 12 місяців (детальніше в додатку Г). Отримана оцінка є середньозваженим середнім значенням цих інтервалів. Відповідь “важко відповісти” також доступна домогосподарствам і такі відповіді виключаються з розрахунку середніх очікувань. Водночас фінансові аналітики надають свої дискретні прогнози інфляції (фактичне число, а не інтервальну оцінку), а їхні очікування є простим середнім цих оцінок. Останнє може призвести до періодичної упередженості, оскільки кількість експертів в опитуванні непостійна.

У таблиці 1 наведено короткий статистичний знімок інфляційних очікувань в Україні. Історично склалося так, що очікування професійних прогнозистів були нижчими, ніж у всіх інших респондентів (середнє значення на 2–4 в. п. нижче, ніж очікування банків, суб'єктів підприємницької діяльності і домогосподарств). Однак вони не дають набагато точніших прогнозів, оскільки похибка прогнозу коливається в різних напрямках (рисунок 3), як і в інших респондентів. Таким чином, RMSE (середнє квадратичне відхилення) очікувань фінансових аналітиків становить 12.0 в. п., що вище за RMSE очікувань домогосподарств і підприємств (11.4 в. п. і 11.3 в. п. відповідно за аналогічний період з липня 2014 р.). Очікування банків показують найгіршу прогностичну силу, з RMSE 13.1 в. п.

Усі очікування мають позитивний перекис, а це означає, що праві хвости розподілу досить

Таблиця 1. Статистичний характер інфляційних очікувань

	Банки		Суб'єкти підприємницької діяльності		Домогосподарства	Фінансові аналітики
	Повний обсяг вибірки	3 липня 2014 року	Повний обсяг вибірки	3 липня 2014 року	Повний обсяг вибірки (з липня 2014 року)	Повний обсяг вибірки (з липня 2014 року)
Підрахунок (квартали або місяці)	38	28	61	27	81	75
Середнє значення, %	10.660	12.120	13.070	13.430	13.780	9.880
Стандарт, відсотковий пункт	4.990	4.880	5.390	6.920	4.680	3.430
Мінімально, відсотковий пункт	3.500	5.800	4.700	5.100	4.510	5.340
25%, відсотковий пункт	6.890	9.150	9.000	7.800	9.790	7.200
Серединне значення, відсотковий пункт	9.920	10.650	12.760	10.000	13.550	8.800
75%, відсотковий пункт	12.000	14.330	15.800	18.650	17.140	12.180
Максимально, відсотковий пункт	24.900	24.900	27.300	27.300	22.890	21.900
Коефіцієнт асиметрії розподілу	1.074	1.045	0.731	0.811	0.108	1.082
Коефіцієнт ексцесу розподілу	0.653	0.556	0.132	-0.699	-1.002	0.852

довгі. Водночас очікування домогосподарств майже симетричні, маючи лише невеликий хвіст розподілу праворуч. Розподіл інфляційних очікувань домогосподарств більш пологий, ніж зазвичай, тоді як усі інші очікування є більш піковими.

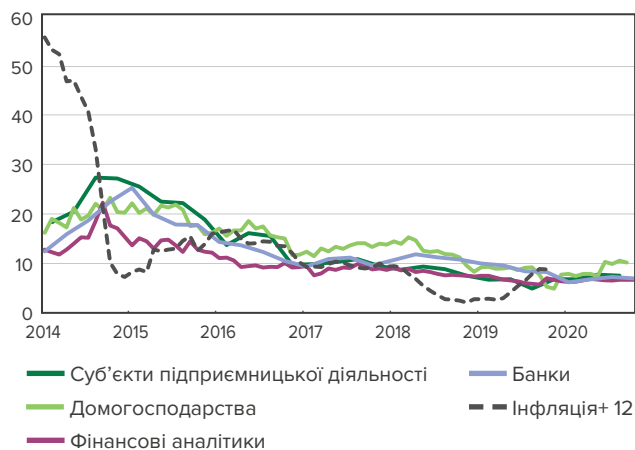


Рисунок 3. Інфляційні очікування на наступні 12 місяців та фактична інфляція (+12 місяців), %

Джерело: Державна служба статистики України, НБУ, GfK Ukraine, Info Sapiens.

3. РОЗРАХУНОК СУКУПНИХ ПОКАЗНИКІВ НОВИН

Контент у сукупності новин пов'язаний переважно з економічними, соціальними та політичними темами. Таким чином, вибірка включає новини, які пов'язані не тільки з розвитком інфляції чи очікуваннями (ціни, пропозиція окремих товарів, тарифи, статистична інформація, прогнози тощо). Щоб зосередитися лише

на чинниках, що визначають інфляційні очікування, ми відкидаємо новини, які не стосуються теми. Для цього застосовуємо два різних підходи до фільтрації новин. По-перше, ми використовуємо словниковий підхід для створення набору показників на основі необробленої кількості новин. По-друге, ми реалізували тематичний аналіз за допомогою латентного розміщення Діріхле (ЛРД) згідно з Blei et al. (2003).

Обидва підходи не враховують емоційного забарвлення контенту новин. Однак це не може бути великою проблемою, оскільки зазвичай новини негативно упереджені – Hester & Gibson (2003) виявили, що економічні новини частіше писалися в негативному тоні, ніж у позитивному. Крім того, вони довели, що негативні новини є значним провісником очікувань споживачів щодо майбутнього економічного розвитку. Damstra & Boukes (2018) добре пояснюють цю негативну упередженість новин, наводячи кілька основних причин:

- вільні ЗМІ відіграють вирішальну роль у нагляді за урядом, тому негативні події отримують більше уваги, а позитивні не задовольняють такої потреби;
- у процесі оцінювання важливості висвітлення інформації стосовно подій реального світу негативність може бути ключовою цінністю, отже, «погана» новина, найімовірніше, буде обрана журналістами;
- негативні події мають сильніший новинний вплив, ніж позитивні.

Крім того, Soroka et al. (2019) визначають негативну тональність новин як суттєву ознаку, тоді як хороші новини, навпаки, можна розглядати як брак новин. Тому для отримання простих показників можна припустити

наявність тенденції негативного впливу новин на сприйняття та очікування.

Звичайно, аналіз емоційного забарвлення повідомлень може бути корисним для визначення впливу новин на економічні очікування. Зокрема, такий підхід може очистити ряд даних від суперечливих подій, які можуть мати протилежні наслідки. Крім того, поділ новин на позитивні та негативні допоміг би дослідити можливу нелінійність впливу різноспрямованих настроїв. Проте для застосування цього підходу важливо створити якісний набір даних для навчання, заснований на чітких правилах, із залученням кількох незалежних експертів. Крім того, аналіз емоційного забарвлення повідомлень потребує значної обчислювальної потужності та часу, що може бути недоцільним на початкових етапах дослідження.

3.1. Словниковий підхід

Словниковий підхід є найпростішим підходом до оцінки впливу новин на різні макроекономічні показники. Ці показники розраховуються як частка статей, пов'язаних із темою, що зазвичай позначається як “частота документів”. Інтуїція, що стоїть за цими показниками, полягає в тому, що чим тривожнішою є тема, тим більше статей буде написано на цю тему – наприклад, під час кризи.

Частота документа (df) це частка документів, які містять певний термін, отримується шляхом ділення кількості документів, що містять цей термін, на загальну кількість документів:

$$df(t, D) = \frac{d}{N}, \quad (1)$$

де N – загальна кількість документів у сукупності новин D , $|\{d \in D : t \in d\}|$ – кількість документів d у яких зустрічається t термін.

Словниковий підхід до визначення показників новин потребує досвіду у виборі відповідних ключових слів. Приміром, щоб визначити, які ціни найбільше хвилюють українців, ми звернулися до споживчого кошика середньостатистичного домогосподарства. Найбільше своїх доходів українці витрачають на їжу. У різні періоди частка витрат на продукти харчування та безалкогольні напої становила 40–60% за період з 2000 по 2020 рік, дещо знизившись в останні роки. Тому важливо відбирати новини, які містять згадки про основні продукти харчування: хліб, м'ясо, молоко, овочі, фрукти тощо.

Ще одна важлива складова витрат домогосподарств – комунальні послуги. Хоча частка цього виду витрат значно нижча, ніж у багатьох інших країнах, тарифи на комунальні послуги важливі для українців і часто використовуються політиками як так званий “політичний футбол”. Тому вони можуть мати помітний вплив на очікування. Найважливішими комунальними послугами для українців є електроенергія та природний газ.

Ціни на пальне також можуть мати значний вплив на формування інфляційних очікувань домогосподарств, хоча не всі люди користуються приватним транспортом. Наприклад, Kilian and Zhou (2020) знайшли кілька епізодів із 1990 року в Сполучених Штатах Америки, коли зростання інфляційних очікувань домогосподарств

майже повністю можна було пояснити підвищенням цін на пальне. З одного боку, це пов'язано з повсюдним поширенням АЗС і табло цін на пальне, що дає змогу легко використовувати їх для щоденного моніторингу цін. З іншого боку, усі добре знають, що пальне є складовою вартості більшості товарів і послуг, явно чи неявно. У цьому випадку ми включаємо новини не тільки про пальне, а й про нафту, яка є визначальною складовою вартості палива.

Як зазначено в праці Coibion and Gorodnichenko (2015b), існує сильна позитивна кореляція між інфляційними очікуваннями та змінами обмінного курсу – особливо для домашніх господарств. Водночас, за логікою, на очікування впливає не тільки курсова динаміка, а й висвітлення цієї теми в ЗМІ.

Крім того, ми проаналізуємо показник новин, пов'язаних із самим словом інфляція, оскільки такі новини часто містять експертні прогнози чи аналізи поточної ситуації. Згідно з Zholud et al. (2019), інфляційні очікування в Україні тісно пов'язані з поточними інфляційними тенденціями, тому мають компонент, орієнтований на майбутнє. Відповідно доцільно перевірити вплив згадок про інфляцію в ЗМІ на формування очікувань.

На рисунку 4 подано отримані показники, розраховані за допомогою словникового підходу. Оскільки найперші дані щодо інфляційних очікувань респондентів датуються 2006 роком, усі часові ряди показників новин починатимуться з 2005 року (один рік назад, щоб оцінити явище запізнення). Однак, урахувавши, що на початку 2000-х років обсяг новин був набагато меншим, це означало, що лише близько 10% статей потрібно було видалити із сукупності новин, і все одно залишається близько 1.8 мільйона статей. Новини, пов'язані з їжею, мають найменшу частку серед обраних тем, тоді як новини про паливо є найважливішими. Також видно, що частота документів з новинами, пов'язаними з комунальними послугами, загалом із часом зменшилася – за винятком 2015 року, коли відбувся значний стрибок важливості цієї теми через підвищення комунальних тарифів до ринкового рівня. Цікаво, що до 2014 року тема валютних курсів згадувалася в новинах частіше, що, ймовірно, пов'язано з більшими негативними наслідками різкої девальвації, яка спостерігалася в Україні тоді після відмови від режиму фіксованого курсу. Для отримання додаткової інформації про показники, створені за допомогою словникового підходу (додаткові матеріали, Appendix E).

Оскільки інфляційні очікування різних респондентів збираються в різні періоди, а не рівномірно протягом місяця, вплив певних короточасних або навіть дискретних новин може бути надзвичайно важливим. Таким чином, деякі новини можуть “жити” лише кілька днів, і через швидку втрату інтересу до теми вплив на частоту щомісячних публікацій може зникнути. Тому місячні показники можуть не відображати реальної динаміки важливості окремих подій, а застосування показників з більшою частотою може пролити світло на це питання. Щоб оцінити цей вплив, ми додатково розрахували аналогічні показники в декадному (10-денному) терміні для кожного місяця – декада є третиною місяця (результати наведено в додаткових матеріалах, Appendix E).

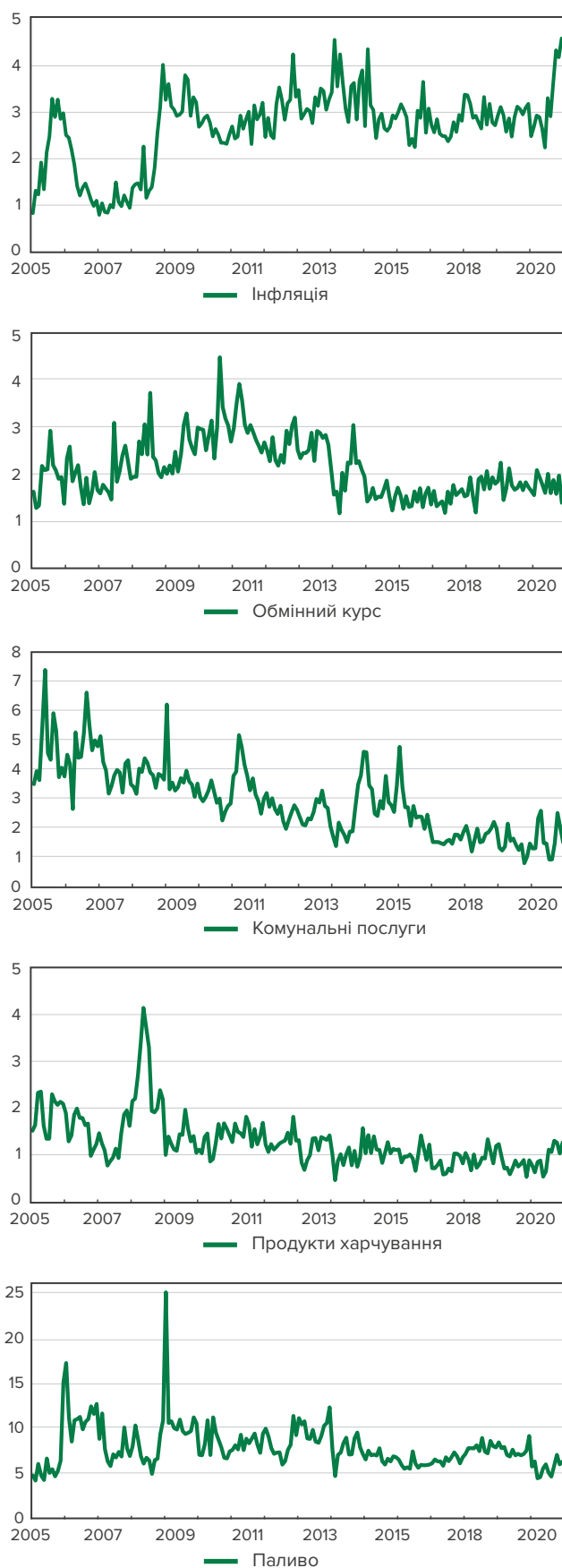


Рисунок 4. Частота документів на теми, які стосуються інфляційних очікувань

Важливо припустити незалежність досліджуваних змінних, про що може свідчити їхня стаціонарність (незмінність). Стаціонарність потрібна для застосування

багатьох статистичних інструментів і процедур в аналізі часових рядів. Дійсно, якщо дані були згенеровані стаціонарним процесом, вони матимуть властивості зразка, згенерованого таким процесом. Відповідно до тесту Дікі – Фуллера, місячний часовий ряд частки тем про комунальні послуги, курс валют та інфляцію в новинах є нестаціонарним. Як наслідок, відповідні показники частки тем комунальних послуг та інфляції в новинах у десятиденному часовому ряді є нестаціонарними. Крім того, автокореляція висока і здається, що немає чіткої сезонності. Тому, щоб позбутися високої автокореляції та зробити весь процес стаціонарним, ми в такий самий спосіб відберемо перші відмінності.

3.2. Підхід щодо машинного навчання без учителя

Одним із важливих недоліків словникового підходу є потреба проведення якісної експертизи та відбору текстів на її основі. Зокрема, стаття може містити ключові слова, але її тема може бути зовсім іншою. Наприклад, слово “паливо” можна віднести до тем, пов’язаних з наукою і технікою або автомобілебудуванням. Рішенням тут є використання алгоритмів тематичного моделювання без учителя. Ці статистичні методи аналізують слова в сукупності текстів і ділять їх на підгрупи, де кожна підгрупа пов’язана з набором ключових слів. Таким чином, модель знаходить поєднання слів, а не окремі. У нашому прикладі з ключовим словом “паливо” виділялися б статті зі словосполученнями “ціна палива” та “ракетне паливо”. Більшість моделей машинного навчання потребують використання частини набору даних, коли спеціально навчені люди класифікують інформацію відповідно до заздалегідь визначеної процедури та позначають ідентифікаторами дані. Однак деякі методи не потребують таких ідентифікованих навчальних зразків. Латентне розміщення Діріхле (ЛРД), представлене Блаєм та іншими у 2003 році, на сьогодні дуже поширений приклад методу моделювання теми, який використовує алгоритм навчання без учителя.

Ми використали надзвичайно ефективну версію ЛРД під назвою LightLDA, яку можна знайти в модулі `nimbusml` Python (Yuan et al., 2015). Ця найсучасніша версія пропонує кілька методів оптимізації та може набагато швидше “натренувати” тематичну модель на великих наборах документів. Наприклад, наша модель створює 100 тем у системі даних із 2 мільйонів новин менш ніж за годину, тоді як використання повного ЛРД у такому масштабі займає кілька днів. На рисунку 6 (Додаток Б) показано розподіл тем, отриманих за допомогою ЛРД. Популярність одних тем змінювалася з часом, а інші залишалися актуальними протягом усього періоду спостереження.

Кількість тем в ЛРД не фіксована і може бути встановлена відповідно до завдання. Ми експериментували з використанням різної кількості тем. Було помічено, що з більшою кількістю тем наші основні результати не змінюються – деякі теми мають дуже схожий зміст і їх потрібно об’єднати в подальшому аналізі. Водночас ускладнюється інтерпретація більшої кількості тем. Із меншою кількістю тем іноді важко розрізнити різні теми, які мають схожі ключові слова. Наприклад, теми, пов’язані з курсами валют, можуть

містити непотрібну інформацію, оскільки деякі статті містять схожі слова, але різний зміст.

На цьому етапі потрібне втручання людини, щоб проаналізувати та позначити теми отриманих сукупностей новин. На рисунку 7 (Додаток Б) наведено граф що показує зв'язок між темами, розподіленими ЛРД. Більшість сукупностей новин належить до такої тематики: політика, міжнародні відносини, парламент та уряд. Водночас деякі теми відрізняються і можуть бути пов'язані з економічними темами, які можуть вплинути на інфляційні очікування. Більшість тем новин належать не до однієї, а до кількох сукупностей новин.

Нам вдалося ідентифікувати групу новин, пов'язану зі зміною валютного курсу, яка містить шість тем, визначених за допомогою ЛРД. Як відомо, певний вплив на формування інфляційних очікувань в Україні має ситуація на валютному ринку. Ми також знайшли групу, пов'язану з товарами, включаючи нафту і газ. Крім того, можна легко визначити теми, що описують ринок електроенергії, бюджет та державний борг. Цікаво, що ЛРД допомогло визначити тему, пов'язану з поширенням коронавірусу, кількість статей щодо якої з очевидних причин зросла з кінця 2019 року. Крім того, ЛРД має чітко визначену тематику періоду війни між росією та Україною з 2014 року та наступних років. Однак ЛРД не об'єднало статті, пов'язані з цінами на продукти харчування, тарифами на комунальні послуги тощо, в окремі впізнавані теми, що можна пояснити їхньою подібною структурою, а також відносно низькою часткою таких новин. Цю проблему збільшення кількості тем не вирішує. На рисунку 8 (Додаток Б) показано "хмари слів" для кількох найбільш актуальних тем.

Популярність окремих тем чітко відповідає історичному розвитку подій. Зокрема, частка статей про обмінний курс гривні, зображених на рисунку 5, зросла у 2008 році, коли гривня стрімко знецінювалася в умовах світової фінансової кризи. Наступний пік припав на 2014–2015 роки, коли через війну між росією та Україною та втрату контролю над частиною території України економіка зазнала значного удару. У цей час також стрімко знецінювався курс гривні. Але з переходом до плаваючого курсу і стабілізацією на валютному ринку інтерес до цієї теми в новинах почав згасати.

Подібна тенденція спостерігалася і щодо новин про газ і нафту. Так, у 2006–2008 роках газове питання було надзвичайно актуальним для України на тлі складних відносин з росією. У 2014 році проблеми з газопостачанням повторилися. Водночас нижчі ціни на

енергоносії сприяли меншому висвітленню цих тем у наступні роки.

Ми визначили показники так само, як і за словникового підходу, використовуючи рівняння 1 частоти документа. Таким чином, місячні показники були розраховані для оцінки довгострокового впливу, тоді як десятиденні показники були розраховані для оцінки короткострокових медіашоків, які можуть бути важливими під час дослідження інфляційних очікувань, але потім згасають протягом місяця. Більш детально зі статистичними характеристиками показників, отриманих розподілом ЛРД, можна ознайомитись у додаткових матеріалах, Appendix F. Відповідно до тесту Дікі – Фуллера, місячні часові ряди для частки енергетичних новин є нестационарними, тоді як частка новин щодо зміни валютного курсу стаціонарна. Декадні часові ряди можна вважати стаціонарними з імовірністю 95%.

4. РЕЗУЛЬТАТИ ОЦІНЮВАННЯ

Як зазначалося в попередніх розділах, інфляційні очікування значною мірою формуються інфляцією попередніх періодів (Zholud et al., 2019; Coibion and Gorodnichenko, 2015b). Тому для аналізу ми використали екстраполяційний підхід до формування інфляційних очікувань (Lines and Westerhoff, 2010):

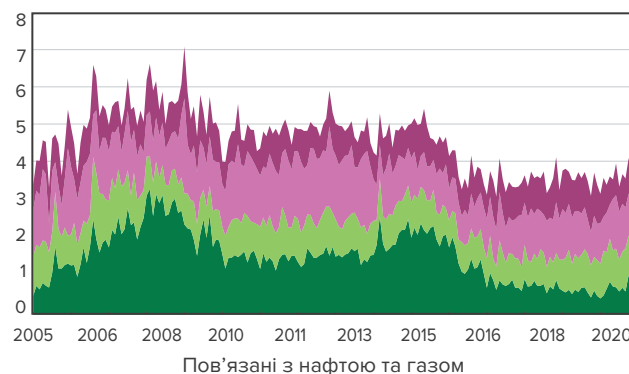
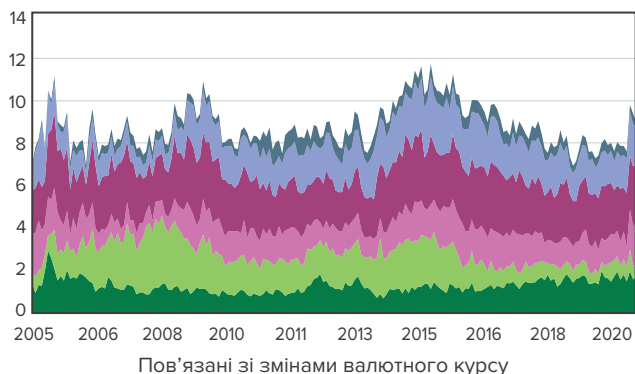
$$E\pi_t = \alpha + \beta\pi_{t-1} + \gamma(\pi_{t-1} - \pi_{t-2}) + \varepsilon, \quad (2)$$

де $E\pi_t$ – очікувана інфляція протягом періоду t , π_{t-1} позначає інфляцію в попередньому періоді, $\pi_{t-1} - \pi_{t-2}$ означає зміну інфляції, α та γ – коефіцієнти регресії, а ε це похибка. Ми використовуємо річну зміну індексу споживчих цін як міру інфляції.

У цьому дослідженні ми припускаємо, що на формування інфляційних очікувань (рівняння 2) впливає медіасередовище, а не фактичні зміни в інфляції:

$$E\pi_t = \alpha + \beta\pi_{t-1} + \delta df_T^m + \varepsilon, \quad (3)$$

де df позначає частоту документа на тему новин m протягом періоду T . T може дорівнювати у разі тестування впливу новин на формування інфляційних очікувань у тому самому місяці, коли проводиться опитування. Однак деякі опитування проводяться на початку місяця, тому ми тестуємо вплив частоти публікацій новин протягом попередніх трьох місяців на формування інфляційних очікувань. Відповідно T може дорівнювати $t-1$, $t-2$ і $t-3$. Ми перевіряємо місячну та декадну періодичність T , оскільки дослідження інфляційних очікувань проводиться не за



Рисунки 5. Частка тем, визначених за допомогою ЛРД, частота документа, %

цілий місяць, а за коротші періоди. Крім того, ці періоди також відрізняються для різних респондентів. Оскільки щоквартальні опитування не проводяться протягом кварталу, ми використовуємо порівняння місяців замість агрегування показників новин на квартальному рівні. Наприклад, опитування банків проводяться в перший місяць кварталу, тому той самий місяць для показника новин використовувався як базовий.

Ми також тестуємо інший варіант екстраполяційних інфляційних очікувань, припускаючи, що очікування респондентів змінюються у відповідь на зміни поточної інфляції. У цьому випадку формула інфляційних очікувань має такий вигляд:

$$E\pi_t - E\pi_{t-1} = \alpha + \gamma(\pi_{t-1} - \pi_{t-2}) + \varepsilon. \quad (4)$$

Ми поширили формулу 4 зі змінами медіасередовища на зміни поточної інфляції:

$$E\pi_t - E\pi_{t-1} = \alpha + \gamma(\pi_{t-1} - \pi_{t-2}) + \eta(df_T^m - df_{T-1}^m) + \varepsilon. \quad (5)$$

У такому разі всі наші змінні стаціонарні, і ми можемо бути впевнені, що їхні властивості не змінюються з часом.

Ми також припускаємо, що вплив визначених показників новин на інфляційні очікування є лінійним, тому для оцінки цього ефекту ми використовуємо регресію МНК (метод найменших квадратів).

Почнемо з аналізу впливу новин на формування інфляційних очікувань за допомогою словникового підходу. У таблиці 2 наведено коефіцієнти та р-значення (в дужках) показників новин, отриманих за допомогою словникового підходу в МНК-регресіях інфляційних очікувань різних груп респондентів. У цій таблиці показано два різних підходи: без перетворень (рівняння 3), з використанням усіх змінних у тому вигляді, як вони були обчислені, і перші різниці всіх змінних, що є екстраполяцією зміни інфляційних очікувань (рівняння 5). Коефіцієнт детермінації R^2 для першого типу зв'язку, як і очікувалося, є набагато вищим, ніж для оцінок першої різниці. Однак відносно низький R^2 для таких досліджень цілком нормальний і типовий для досліджень людської поведінки (King, 1986).

Як впливає з таблиці 2 (Додаток А), усі типи інфляційних очікувань залежать від поточних інфляційних тенденцій, оскільки коефіцієнти є статистично значущими. Водночас лише банки та суб'єкти підприємницької діяльності пов'язують зміни в інфляційних очікуваннях з останніми змінами інфляції, тоді як зв'язок між змінами в очікуваннях домогосподарств і фінансових аналітиків з останньою динамікою інфляції є незначним. Це узгоджується з думкою, що добре закріплені довгострокові інфляційні очікування не повинні змінюватися у відповідь на новини про макроекономічні показники, зокрема про інфляцію (Galati et al., 2011). Однак стверджувати про закріплені інфляційні очікування поки зарано, ураховуючи різницю між інфляційними цільовими показниками центрального банку та інфляційними очікуваннями. Тому в цьому випадку результат може бути наслідком інформаційної жорсткості.

Інфляційні очікування банків практично не залежать від поточного медіасередовища щодо інфляції. Більшість показників не є статистично значущими або суперечать економічній логіці. Наприклад, інфляційні очікування

банків негативно корелюють з продовольчими новинами з ймовірністю 90%. Тобто чим більше про цю тему говорять у ЗМІ, тим швидше вона знижує інфляційні очікування банків. Це можна пояснити тональністю та змістом новин. Однак без детальнішого вивчення змісту відповідних новин це неможливо визначити.

Водночас характерно, що банки змінюють свої інфляційні очікування під впливом змін інформаційного середовища навколо тарифів на комунальні послуги та цін на паливо, а також новин про інфляцію за попередні періоди. Досить значне відставання в два-три місяці можна пояснити часом, потрібним для підготовки макроекономічних прогнозів, які є основою для відповідей в опитуванні.

Подібно до банків на очікування бізнесу можуть суттєво вплинути новини про попередні тенденції інфляції та тарифи на комунальні послуги. Однак бізнес найбільш чутливий до новин про продукти харчування. Це можна пояснити високою часткою серед опитаних підприємств сільського господарства, харчової промисловості, роздрібною та оптовою торгівлі (пов'язаною з продуктами харчування), що також відповідає структурі економіки України. Продовольчі новини також є важливим чинником під час оцінювання зміни інфляційних очікувань суб'єктів підприємницької діяльності.

Очікування домогосподарств можуть бути найбільш чутливими до кількості новин про тарифи на комунальні послуги у звітному періоді та попередньому кварталі. Насамперед це можна пояснити високою значущістю тарифів на комунальні послуги для українських домогосподарств. Таким чином, значна частка тарифів регулюється державою або місцевою владою, а зміни тарифів викликають суттєву негативну реакцію суспільства. Частка тарифів на комунальні послуги в індексі споживчих цін є відносно низькою, що значною мірою пов'язано з негрошовими субсидіями, які діяли в попередні періоди. Однак, незважаючи на це, середні українські тарифи на комунальні послуги є одними з найважливіших тем, пов'язаних з інфляцією, що підтверджують, зокрема, результати нашого аналізу. Домогосподарства також трохи чутливі до інформації про продукти харчування в попередні періоди. Цікаво, що ці результати не підтверджуються змінами в інфляційних очікуваннях домогосподарств. Таким чином, громадяни змінюють свої оцінки майбутньої інфляції під впливом змін в інформаційному полі про курс за попередні три місяці, тоді як зміна важливості інших тем мало впливає на динаміку очікувань.

Очікуванням професійних прогнозистів найкраще відповідає інформація про тарифи на комунальні послуги у звітному та попередніх місяцях, а також курс валют за попередні три місяці. Водночас фінансові аналітики реагують як на кількість інформації, так і на її зміну. Це може відображати різні підходи таких аналітиків до прогнозування. Найбільше на перегляд прогнозів, як завжди, впливають зміни курсу та очікувані зміни тарифів на комунальні послуги.

Ми також вирішили перевірити гіпотезу про те, що більш короткострокові тенденції в медіасередовищі можуть краще пояснити процес формування інфляційних очікувань у різних респондентів. Це узгоджується з тим фактом, що більшість опитувань проводяться за короткий період, ніж місяць. Для цього

ми використовуємо декадні позначення частоти згадок певних тем. Потрясіння в новинах, які тривають кілька днів, можуть зійти нанівець протягом місяця через швидку втрату інтересу до теми, а тому місячні показники можуть не відображати реальної динаміки важливості окремих подій. Таким чином, важливо застосовувати показники з більшою частотою. Переходячи до рівня декад ми отримуємо змішану частоту в МНК, тому, щоб перейти до однієї частоти, ми використовували відповідність за місяцями. Таким чином, ми порівнюємо показники новин окремо за першу, другу та третю декаду звітного місяця з інфляційними очікуваннями за той самий звітний місяць. Процедура було повторено для окремих показників новин за три декади попереднього місяця щодо інфляційних очікувань звітного місяця, оскільки респонденти також реагували на динаміку медіасередовища в попередні періоди.

У таблиці 3 (Додаток А) наведено результати оцінок МНК щодо впливу десятиденних показників новин на формування інфляційних очікувань. Як і в попередньому випадку, ми додаємо останні доступні дані про інфляцію, які публікуються із запізненням, тому використовуємо фактичну інфляцію за період t-1.

Тут ми маємо кілька цікавих результатів, які відрізняються від наших щомісячних оцінок. Наприклад, банки можуть бути чутливими до частоти документа про обмінний курс у всіх декадах попереднього місяця, тоді як місячні показники не відображають цього зв'язку. Водночас новини про продукти харчування можуть бути більш важливими в останню декаду попереднього місяця, хоча місячні показники свідчать про значущість для поточного місяця. Суб'єкти підприємницької діяльності більше реагують на новини про комунальні послуги та паливо в першій декаді звітного місяця. Як і місячні показники, декадні показники тарифів на комунальні послуги можуть впливати на формування інфляційних очікувань домогосподарств. Тим часом найважливішими є перші декади звітності та попередні місяці. Очікування фінансових аналітиків також виявилися найбільш залежними від частоти новин у перші декади звітності та попередні місяці. Однак, крім комунальних послуг, вони можуть стежити за новинами про інфляцію (офіційні дані публікуються лише в першій декаді місяця) і про паливо.

Інша важлива думка стосується того, що інформація про очікування банків і підприємств збирається раз на квартал. Тому період оцінки впливу новин на інфляційні очікування було збільшено шляхом застосування тримісячної ковзної середньої. Це особливо важливо, ураховуючи, що коефіцієнти для місячних показників дуже мінливі і можуть навіть змінювати знак залежно від застосованого запізнення. У таблиці 4 (Додаток А) наведено результати МНК-оцінки впливу кварталних показників новин (тримісячна ковзна середня) на формування інфляційних очікувань.

Як бачимо, гіпотеза про те, що банки та підприємства керуються більш тривалими трендами, здебільшого не підтверджується. Водночас довгострокова зміна інформаційного простору щодо інфляції та обмінного курсу пов'язана зі зміною інфляційних очікувань банків, а зміна обсягу новин про продукти харчування впливає на інфляційні очікування суб'єктів підприємницької діяльності. Однак в обох випадках цей вплив обмежується одним-двома кварталами.

Ми повторили аналогічну процедуру, щоб виявити вплив показників, визначених за допомогою ЛРД, на формування інфляційних очікувань. У таблиці 5 (Додаток А) наведено коефіцієнти та r-значення (в дужках) показників новин, отриманих ЛРД в МНК-регресіях інфляційних очікувань різних груп респондентів. Подібно до простих показників ми протестували два різних підходи: без перетворень, з використанням усіх змінних у тому вигляді, як вони були обчислені, і першу різницю всіх змінних, що представляє екстраполяцію зміни інфляційних очікувань.

За результатами регресій ми спостерігаємо слабку відповідність між новинами про тарифи на енергоносії та комунальні послуги, які визначило ЛРД, та формуванням інфляційних очікувань. Інфляційні очікування домогосподарств, як і зміни ділових очікувань, демонструють значущість у звітному місяці на рівні 5% та 10% відповідно, але знак коефіцієнтів цих змінних суперечить економічній логіці, що може бути пов'язано або з емоційним забарвленням, або ж впливом збігу обставин.

Натомість дещо інша ситуація з новинами про зміну курсів валют, отриманими за допомогою ЛРД. Частота таких новин у звітному періоді була суттєвою для формування очікувань бізнесу, домогосподарств та фінансових аналітиків. Для домогосподарств і фінансових аналітиків ці показники також були важливими в останні місяці. Фінансові аналітики та домогосподарства також були чутливими до зміни частоти обмінних курсів валют. Однак домогосподарства змінюють свої очікування у відповідь на останні події, тоді як фінансові аналітики реагують на більш тривалий період.

Подібно до простих показників ми визначили короткострокові спотові ефекти на формування інфляційних очікувань шляхом оцінювання десятиденних показників. У таблиці 6 (Додаток А) наведено результати цієї оцінки.

Показово, що для деяких груп респондентів спостерігається чіткий зв'язок між показниками в періоді проведення опитувань. Наприклад, опитування банків зазвичай проводяться на початку місяця, а іноді навіть охоплюють останній тиждень попереднього місяця. Вагомими виявилися новини щодо енергоресурсів в останній декаді попереднього місяця та в першій декаді звітного місяця. Подібна ситуація спостерігається стосовно суб'єктів підприємницької діяльності та домогосподарств. Водночас знак коефіцієнтів потребує додаткового вивчення стосовно емоційного забарвлення повідомлень. Інфляційні очікування суб'єктів підприємницької діяльності формуються під впливом новин про курси валют першої декади звітного місяця, а решта респондентів стежить за новинами за попередні періоди.

Ми повторили таку саму процедуру для визначення довгострокового впливу новин на формування інфляційних очікувань, використовуючи поточне тримісячне середнє значення для банків і компаній. У таблиці 7 (Додаток А) викладено основні результати оцінок. Однак результати свідчать про те, що довгострокового впливу новин на формування інфляційних очікувань немає. Лише очікування суб'єктів підприємницької діяльності мають суттєвий зв'язок зі

зміною частоти новин про курси валют у поточному кварталі.

Таким чином, формування інфляційних очікувань у різних груп респондентів може залежати від медійного середовища, а саме як від обсягу опублікованих статей, так і від зміни цього показника. Важливо зазначити, що різні групи респондентів можуть спиратися на різні теми та різні періоди, коли оцінюють майбутню інфляцію. Також можна побачити, що останні новини, опубліковані протягом місяця і навіть декади, що передували опитуванню, здебільшого важливіші для формування інфляційних очікувань, ніж старі новини. Це може бути, з-поміж іншого, важливим для комунікаційної політики центрального банку.

5. ВИСНОВКИ

У цій праці ми проаналізували роль новин у формуванні інфляційних очікувань різних типів респондентів в Україні за допомогою текстових даних. Ми зібрали сукупність новин із чотирьох українських інтернет-видань, які входять до списку найпопулярніших онлайн-медіа України та які переважно мають економічну спрямованість. Використовуючи оброблення природної мови та методи машинного навчання, ми очистили й перетворили текстові дані в кількісні показники на основі новин, які відображають теми новин, пов'язані з інфляцією та інфляційними очікуваннями.

Ми застосували два різних підходи для фільтрації новин: словниковий підхід та латентне розміщення Діріхле (ЛРД). Обидва підходи не враховують емоційного забарвлення вмісту новин, яке ми залишаємо для майбутніх досліджень. Ми обчислили показники всіх новин як «частоту документа», інтуїтивно припустивши, що чим тривожнішою є тема, тим більше статей буде написано на цю тему.

Ми припустили, що вплив отриманих показників новин на інфляційні очікування є лінійним, і оцінили цей ефект за допомогою регресії МНК. Ми протестували вплив на рівень інфляційних очікувань, а також на їх зміну. Ми знайшли докази того, що різні теми новин можуть по-різному впливати на інфляційні очікування різних груп. Наприклад, виявлено тісний зв'язок між інфляційними очікуваннями домогосподарств і фінансових аналітиків і новинами про комунальні послуги, тоді як суб'єкти підприємницької діяльності чутливі до новин про продукти харчування. Крім того,

фінансові аналітики та домогосподарства також були чутливими до рівнів і змін частоти обмінних курсів, як показує ЛРД.

Ми також перевірили гіпотезу про те, що більш короткострокові тренди в медіасередовищі можуть краще пояснити процес формування інфляційних очікувань різних респондентів, оскільки періодичність документування може змінюватися протягом місяця, а вплив короткострокових новин може зникати. Ми довели, що для деяких груп респондентів є чіткий зв'язок з показниками в періоди проведення опитувань. Ми також продемонстрували, що останні новини здебільшого важливіші для формування інфляційних очікувань, ніж старі новини.

Як наслідок, формування інфляційних очікувань різних груп респондентів може залежати від медійного середовища, а саме обсягу опублікованих статей та зміни цього показника. Різні групи респондентів спираються на різні теми та різні періоди, коли оцінюють майбутню інфляцію. Ми також виявили, що деякі події суперечать економічній логіці, це може стати темою майбутніх досліджень. Зокрема, важливим питанням є вплив показників новин у різні періоди (під час стабільної, прискореної інфляції чи дезінфляції). Інші питання дослідження можуть включати оцінку тональності новин, зв'язок показників новин з іншими макроекономічними показниками, а також прогностичну силу таких показників. Іншим важливим питанням може бути дослідження нелінійних взаємозв'язків. Зокрема, вплив новин може відрізнитися залежно від рівня поточної інфляції (наприклад, учасники економічної діяльності можуть приділяти більше уваги новинам, коли інфляція висока, і навпаки) та режиму монетарної політики.

Ці результати доповнюють попередні дослідження щодо формування раціональних інфляційних очікувань. Іншими словами, загальний рівень інфляційних очікувань, як правило, визначається інфляцією попередніх періодів, а невеликі коливання цілком можна пояснити іншими факторами, зокрема впливом медіасередовища. Результати цього дослідження можуть допомогти зрозуміти інфляційні очікування, особливо тому, що закріплення інфляційних очікувань залишається ключовим викликом для центральних банків. Це може, з-поміж іншого, бути важливим для комунікаційної політики центрального банку та допомогти йому сформулювати чіткі та ефективні повідомлення, а також розробити оптимальну політику.

ЛІТЕРАТУРА

- Angelico, C., Marcucci, J., Miccoli, M., Quarta, F. (2021). Can we measure inflation expectations using Twitter? Bank of Italy Working Papers, 1318. Rome: Bank of Italy. Retrieved from https://www.bancaditalia.it/pubblicazioni/temi-discussione/2021/2021-1318/en_tema_1318.pdf
- Azqueta-Gavaldón, A. (2017). Developing news-based Economic Policy Uncertainty index with unsupervised machine learning. *Economics Letters*, 158, 47–50. <https://doi.org/10.1016/j.econlet.2017.06.032>
- Baker, S. R., Bloom, N., Davis, S. J. (2016). Measuring economic policy uncertainty. *The Quarterly Journal of Economics*, 131(4), 1593–1636. <https://doi.org/10.1093/qje/qjw024>
- Bauer, M. D. (2015). Inflation expectations and the news. *International Journal of Central Banking*, 11(2), 1-40. Retrieved from <https://www.ijcb.org/journal/ijcb15q2a1.pdf>
- Blei, D. M., Ng, A. Y., Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022. Retrieved from <https://dl.acm.org/doi/10.5555/944919.944937>
- Carroll, C. D. (2003). Macroeconomic expectations of households and professional forecasters. *The Quarterly Journal of Economics*, 118(1), 269–298. <https://doi.org/10.1162/00335530360535207>
- Coibion, O., Gorodnichenko, Y. (2012). What can survey forecasts tell us about information rigidities? *Journal of Political Economy*, 120(1), 116–159. <https://doi.org/10.1086/665662>
- Coibion, O., Gorodnichenko, Y. (2015a). Information rigidity and the expectations formation process: A simple framework and new facts. *American Economic Review*, 105(8), 2644–2678. <https://doi.org/10.1257/aer.20110306>
- Coibion, O., Gorodnichenko, Y. (2015b). Inflation expectations in Ukraine: A long path to anchoring? *Visnyk of the National Bank of Ukraine*, 233, 6–23. <https://doi.org/10.26531/vnbn2015.233.006>
- Coibion, O., Gorodnichenko, Y., Weber, M. (2019). Monetary Policy Communications and their Effects on Household Inflation Expectations. NBER Working Paper Series, 25482. Cambridge: National Bureau of Economic Research. <https://doi.org/10.3386/w25482>
- D'Acunzio, F., Malmendier, U., Ospina, J., Weber, M. (2019). Exposure to daily price changes and inflation expectations. CESifo Working Paper, 7798. Munich: CESifo. Retrieved from <https://www.cesifo.org/en/publications/2019/working-paper/exposure-daily-price-changes-and-inflation-expectations>
- Damstra, A., Boukes, M. (2018). The economy, the news, and the public: A longitudinal study of the impact of economic news on economic evaluations and expectations. *Communication Research*, 48(1), 26–50. <https://doi.org/10.1177/0093650217750971>
- Devlin, J., Chang, M.-W., Lee, K., Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1, 4171–4186. <http://doi.org/10.18653/v1/N19-1423>
- Dräger, L., Lamla, M. J. (2017). Imperfect information and consumer inflation expectations: Evidence from microdata. *Oxford Bulletin of Economics and Statistics*, 79(6), 933–968. <https://doi.org/10.1111/obes.12189>
- Galati, G., Heemeijer, P., Moessner, R. (2011). How do inflation expectations form? New insights from a high-frequency survey. BIS Working Papers, 349. Basel: Bank for International Settlements.
- Garcia, J. A., Werner, S. (2018). Inflation news and Euro area inflation expectations. IMF Working Papers, 167. Washington: International Monetary Fund. <https://doi.org/10.5089/9781484363010.001>
- Hester, J. B., Gibson, R. (2003). The economy and second-level agenda setting: A Time-series analysis of economic news and public opinion about the economy. *Journalism & Mass Communication Quarterly*, 80(1), 73–90. <https://doi.org/10.1177/107769900308000106>
- Hutto, C. J., Gilbert, E. (2014). VADER: A parsimonious rule-based model for sentiment analysis of social media text. Proceedings of the International AAAI Conference on Web and Social Media, 8(1), 216–225. <https://doi.org/10.1609/icwsm.v8i1.14550>
- Kilian, L., Zhou, X. (2020). Oil prices, gasoline prices and inflation expectations: A new model and new facts. CEPR Discussion Paper, 15168. London: Centre for Economic Policy Research. Retrieved from <https://cepr.org/publications/dp15168>
- King, G. (1986). How not to lie with statistics: avoiding common mistakes in quantitative political science. *American Journal of Political Science*, 30, 666. <https://doi.org/10.2307/2111095>
- Korobov, M. (2015). Morphological analyzer and generator for Russian and Ukrainian languages. In: Khachay, M., Konstantinova, N., Panchenko, A., Ignatov, D., Labunets, V. (eds) *Analysis of Images, Social Networks and Texts*. AIST 2015. Communications in Computer and Information Science, 542. Springer. https://doi.org/10.1007/978-3-319-26123-2_31
- Larsen, V. H., Thorsrud, L. A., Zhulanova, J. (2021). News-driven inflation expectations and information rigidities. *Journal of Monetary Economics*, 117, 507–520. <https://doi.org/10.1016/j.jmoneco.2020.03.004>
- Lines, M., Westerhoff, F. H. (2010). Inflation expectations and macroeconomic dynamics: the case of rational versus extrapolative expectations. *Journal of Economic Dynamics and Control*, 34(2), 246–257. <https://doi.org/10.1016/j.jedc.2009.09.004>
- Maćkowiak, B., Wiederholt, M. (2009). Optimal sticky prices under rational inattention. *The American Economic Review*, 99 (3), 769-803. <https://doi.org/10.1257/aer.99.3.769>
- Mankiw, N. G., Reis, R.F., Wolfers, J. (2003). Disagreement about inflation expectations. NBER Macroeconomics Annual,

18, 209–248. Cambridge: National Bureau of Economic Research. <https://doi.org/10.1086/ma.18.3585256>

Mazumder, S. (2021). The reaction of inflation forecasts to news about the Fed. *Economic Modelling*, 94, 256–264. <https://doi.org/10.1016/j.econmod.2020.09.026>

Nautz, D., Pagenhardt, L., Strohsal, T. (2017). The (de-) anchoring of inflation expectations: New evidence from the euro area. *The North American Journal of Economics and Finance*, 40, 103–115. <https://doi.org/10.1016/j.najef.2017.02.002>

Pfajfar, D., Santoro, E. (2013). News on inflation and the epidemiology of inflation expectations. *Journal of Money, Credit and Banking*, 45(6), 1045–1067. <https://doi.org/10.1111/jmcb.12043>

Rozovskaya, A., Roth, D. (2019). Grammar error correction in morphologically rich languages: The case of russian. *Transactions of the Association for Computational Linguistics*, 7, 1–17. https://doi.org/10.1162/tacl_a_00251

Sims, C. (2009). Inflation expectations, uncertainty and monetary policy. *BIS Working Papers*, 275. Basel: Bank for International Settlements. Retrieved from <https://www.bis.org/publ/work275.htm>

Soroka, S., Fournier, P., Nir, L. (2019). Cross-national evidence of a negativity bias in psychophysiological reactions to news. *Proceedings of the National Academy of*

Sciences of the United States of America, 116(38), 18888–18892. <https://doi.org/10.1073/pnas.1908369116>

Taboada, M., Brooke, J., Tofiloski, M., Voll, K., Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2), 267–307. https://doi.org/10.1162/COLI_a_00049

Tobback, E., Naudts, H., Daelemans, W., Junqué de Fortuny, E., Martens, D. (2016). Belgian economic policy uncertainty index: Improvement through text mining. *International Journal of Forecasting*, 34(2), 355–365. <https://doi.org/10.1016/j.ijforecast.2016.08.006>

Woodford, M. (2004). Inflation targeting and optimal monetary policy. *Federal Reserve Bank of St. Louis Review*, 86(4), 15-41. St. Louis: The Federal Reserve Bank of St. Louis. Retrieved from <https://files.stlouisfed.org/files/htdocs/publications/review/04/07/Woodford.pdf>

Yuan, J., Gao, F., Ho, Q., Dai, W., Wei, J., Zheng, X., Xing, E.P., Liu, T., Ma, W. (2015). LightLDA: Big topic models on modest computer clusters. *WWW'15: Proceedings of the 24th International Conference on World Wide Web*, 1351–1361. <https://doi.org/10.1145/2736277.2741115>

Zholud, O., Lepushynskiy, V., Nikolaychuk, S. (2019). The Effectiveness of the Monetary Transmission Mechanism in Ukraine since the Transition to Inflation Targeting. *Visnyk of the National Bank of Ukraine*, 247, 19–37. <https://doi.org/10.26531/vnbu2019.247.02>

ДОДАТОК А. ТАБЛИЦІ

Таблиця 2. Зв'язок між щомісячними показниками новин та інфляційними очікуваннями

Респонденти	Змінні	Без трансформацій					Змінні	1-ша різниця				
		Інфляція	Курс валют	Комуніальні послуги	Їжа	Паливо		Інфляція	Курс валют	Комуніальні послуги	Їжа	Паливо
Банки	π_{t-1}	0.2949*** (0.000)	0.2890*** (0.000)	0.2994*** (0.000)	0.2970*** (0.000)	0.2950*** (0.000)	$\pi_{t-1} - \pi_{t-2}$	0.1047** (0.022)	0.1545*** (0.005)	0.1642*** (0.003)	0.1658*** (0.003)	0.1737*** (0.001)
	df_t^m	-1.2029 (0.194)	0.9028 (0.510)	-0.7928 (0.356)	-4.0589* (0.099)	0.0736 (0.824)	$df_t^m - df_{t-1}^m$	0.6139 (0.379)	1.5717 (0.189)	0.096 (0.903)	1.6328 (0.460)	0.3832 (0.220)
	df_{t-1}^m	1.1547 (0.360)	-0.7359 (0.662)	2.2787* (0.099)	2.9647 (0.104)	-0.3926 (0.346)	$df_{t-1}^m - df_{t-2}^m$	1.5263 (0.123)	-2.2497 (0.153)	1.2997 (0.306)	0.8267 (0.629)	-0.1084 (0.772)
	df_{t-2}^m	1.3360 (0.153)	-0.7157 (0.591)	-1.7610 (0.143)	-0.4362 (0.840)	-0.7258 (0.113)	$df_{t-2}^m - df_{t-3}^m$	1.614** (0.031)	0.5906 (0.624)	-2.2488** (0.037)	-3.6356* (0.065)	-0.9720** (0.026)
	df_{t-3}^m	-1.0740 (0.264)	-0.6057 (0.675)	0.7886 (0.345)	-1.4298 (0.534)	0.7137* (0.100)	$df_{t-3}^m - df_{t-4}^m$	-1.037 (0.158)	-0.5096 (0.689)	0.9001 (0.246)	2.9091 (0.167)	0.8387** (0.033)
	C	6.4354* (0.093)	9.5205*** (0.000)	5.7224*** (0.000)	10.1442*** (0.001)	9.7139*** (0.000)	C	-8.3529*** (0.006)	0.9782 (0.546)	-0.4893 (0.664)	-2.2835 (0.388)	-1.1227 (0.573)
	R ²	0.8430	0.8250	0.8400	0.8370	0.8450	R ²	0.5700	0.3340	0.3720	0.3400	0.4130
Суб'єкти підприємницької діяльності	π_{t-1}	0.3452*** (0.000)	0.3698*** (0.000)	0.3173*** (0.000)	7.0696*** (0.000)	0.3552*** (0.000)	$\pi_{t-1} - \pi_{t-2}$	0.1434*** (0.003)	0.1344*** (0.004)	0.1467*** (0.002)	0.1326*** (0.003)	0.1400*** (0.003)
	df_t^m	-0.1827 (0.843)	1.2093 (0.245)	-0.2132 (0.768)	0.3255*** (0.000)	0.2268 (0.451)	$df_t^m - df_{t-1}^m$	1.0633 (0.122)	0.9183 (0.232)	-0.8341 (0.153)	0.9607 (0.277)	0.0557 (0.799)
	df_{t-1}^m	-2.0502* (0.073)	-0.5729 (0.672)	1.5272** (0.030)	3.3463*** (0.005)	0.2061 (0.293)	$df_{t-1}^m - df_{t-2}^m$	0.3751 (0.645)	0.0937 (0.926)	1.1851** (0.037)	-0.5685 (0.659)	0.0689 (0.631)
	df_{t-2}^m	3.1070** (0.026)	0.0296 (0.981)	-0.3270 (0.720)	-5.3502*** (0.002)	0.0973 (0.767)	$df_{t-2}^m - df_{t-3}^m$	-0.3006 (0.768)	-0.3382 (0.720)	-0.2001 (0.787)	2.7441* (0.054)	0.0326 (0.910)
	df_{t-3}^m	-0.6576 (0.541)	0.9159 (0.384)	0.3750 (0.584)	3.6920** (0.050)	-0.1574 (0.582)	$df_{t-3}^m - df_{t-4}^m$	-0.6195 (0.442)	-0.4283 (0.576)	-0.2405 (0.664)	-2.7642** (0.012)	-0.3174 (0.203)
	C	8.1511*** (0.000)	4.9233*** (0.010)	5.0653*** (0.000)	-0.1370 (0.924)	5.5163** (0.020)	C	-1.5599 (0.178)	-0.6747 (0.578)	-0.0132 (0.988)	-0.7796 (0.349)	1.1141 (0.489)
	R ²	0.6960	0.6850	0.7460	0.7360	0.6870	R ²	0.2270	0.1790	0.2190	0.2740	0.1880
Домогосподарства	π_{t-1}	0.2314*** (0.000)	0.2412*** (0.000)	0.1257*** (0.000)	0.1901*** (0.000)	0.2328*** (0.000)	$\pi_{t-1} - \pi_{t-2}$	0.0403 (0.417)	0.0368 (0.440)	0.0538 (0.795)	0.0486 (0.315)	0.0363 (0.456)
	df_t^m	0.3198 (0.708)	0.7091 (0.609)	1.2605** (0.034)	2.2806 (0.146)	0.2741 (0.536)	$df_t^m - df_{t-1}^m$	0.6496 (0.106)	0.9103 (0.155)	-0.3622 (0.268)	0.9230 (0.239)	0.2290 (0.283)
	df_{t-1}^m	-0.1090 (0.905)	0.4006 (0.773)	0.5428 (0.469)	1.8611 (0.253)	0.1923 (0.691)	$df_{t-1}^m - df_{t-2}^m$	-0.5920 (0.169)	-0.5324 (0.406)	0.6644 (0.125)	-0.3849 (0.632)	0.0145 (0.951)
	df_{t-2}^m	0.9379 (0.320)	-0.4846 (0.725)	-0.1165 (0.876)	2.8517* (0.082)	0.3577 (0.463)	$df_{t-2}^m - df_{t-3}^m$	0.6681 (0.136)	-0.7833 (0.223)	-0.6464 (0.135)	1.3106 (0.106)	0.1944 (0.401)
	df_{t-3}^m	1.1311 (0.209)	1.0076 (0.461)	1.5535** (0.011)	1.7897 (0.280)	-0.0602 (0.890)	$df_{t-3}^m - df_{t-4}^m$	-0.0787 (0.861)	1.2966** (0.040)	0.3707 (0.261)	-1.3967* (0.081)	-0.3634* (0.080)

Таблиця 2 (закінчення). Зв'язок між щомісячними показниками новин та інфляційними очікуваннями

Респонденти	Змінні	Без трансформацій					Змінні	1-ша різниця				
		Інфляція	Курс валют	Комунальні послуги	Їжа	Паливо		Інфляція	Курс валют	Комунальні послуги	Їжа	Паливо
Домогосподарства	C	3.3627 (0.305)	7.1208** (0.011)	4.9512*** (0.000)	2.2403 (0.289)	4.7291* (0.092)	C	-1.9764 (0.234)	-1.5633 (0.198)	-0.1267 (0.795)	-0.5184 (0.607)	-0.5716 (0.672)
	R ²	0.626	0.6100	0.7390	0.6690	0.6230	R ²	0.1070	0.1070	0.0590	0.0900	0.0680
Фінансові аналітики	π_{t-1}	0.1624*** (0.000)	0.1704*** (0.000)	2.1542*** (0.000)	0.1344*** (0.000)	0.1667*** (0.000)	π_{t-1} π_{t-2}	-0.0152 (0.696)	-0.0159 (0.668)	-0.0117 (0.752)	-0.0030 (0.940)	-0.0024 (0.953)
	df ^m _t	0.1562 (0.794)	0.0426 (0.967)	0.0672*** (0.000)	2.1366* (0.086)	0.1165 (0.729)	df ^m _t df ^m _{t-1}	0.5383* (0.100)	0.3330 (0.520)	0.3792 (0.148)	0.4323 (0.538)	0.0671 (0.708)
	df ^m _{t-1}	0.9969 (0.133)	-0.1376 (0.892)	1.0333*** (0.004)	1.7494 (0.156)	-0.1388 (0.728)	df ^m _{t-1} df ^m _{t-2}	0.6199* (0.086)	-0.8697* (0.092)	-0.4401 (0.215)	-0.3148 (0.646)	-0.0930 (0.667)
	df ^m _{t-2}	0.7433 (0.257)	-0.297 (0.765)	0.0352 (0.939)	1.4050 (0.259)	0.2002 (0.628)	df ^m _{t-2} df ^m _{t-3}	-0.8320** (0.021)	0.0794 (0.875)	0.8993*** (0.010)	-0.2915 (0.672)	0.1157 (0.595)
	df ^m _{t-3}	1.2712* (0.052)	2.3240** (0.022)	0.7893* (0.077)	1.3447 (0.295)	0.1429 (0.677)	df ^m _{t-3} df ^m _{t-4}	0.4078 (0.284)	1.5838*** (0.002)	-0.6829** (0.013)	0.6863 (0.333)	0.0082 (0.964)
	C	-2.1422 (0.355)	3.7398* (0.058)	1.2051*** (0.001)	1.1574 (0.491)	4.8133** (0.033)	C	-2.2581* (0.097)	-1.9795** (0.039)	-0.4214 (0.296)	-0.5979 (0.513)	-0.7690 (0.523)
	R ²	0.6750	0.6290	0.8380	0.6600	0.5980	R ²	0.1640	0.1750	0.1480	0.0230	0.0100

Примітки: У таблиці наведено результати регресії МНК, де інфляційні очікування є залежною змінною. Індикатор часу T частоти документа встановлено для t, t-1, t-2 і t-3. Перші цифри в клітинках означають коефіцієнти, а р-значення зазначені в дужках *р<0.1; **р<0.05; ***р<0.01.

Таблиця 3. Зв'язок між показниками десятиденних новин та інфляційними очікуваннями

Респонденти	Змінні	Інфляція		Курс валют		Комунальні послуги		Їжа		Паливо	
		Поточн.	Попер.	Поточн.	Попер.	Поточн.	Попер.	Поточн.	Попер.	Поточн.	Попер.
Банки	π_{t-1}	0.3025*** (0.000)	0.3007*** (0.000)	0.2962** (0.000)	0.2924*** (0.000)	0.2924*** (0.000)	0.2767*** (0.000)	0.2982*** (0.000)	0.3023*** (0.000)	0.2959*** (0.000)	0.2892*** (0.000)
	df ^m I	-0.2913 (0.719)	1.0018 (0.272)	-0.5795 (0.593)	-1.7788** (0.030)	0.3746 (0.652)	0.3437 (0.609)	-0.3568 (0.783)	-0.766 (0.452)	-0.5005 (0.287)	0.0193 (0.961)
	df ^m II	-0.7864 (0.142)	0.8323 (0.346)	-0.3309 (0.742)	3.6288*** (0.007)	0.2311 (0.732)	0.8281 (0.291)	-1.3411 (0.334)	-0.2261 (0.849)	-0.1320 (0.648)	-0.3379 (0.225)
	df ^m III	0.5547 (0.255)	-0.2232 (0.778)	0.4306 (0.680)	-2.2436** (0.012)	-0.0498 (0.932)	-0.2502 (0.738)	-0.7180 (0.637)	1.8529* (0.098)	0.1892 (0.568)	-0.0686 (0.799)
	C	8.6576*** (0.002)	2.2805 (0.494)	8.2388*** (0.000)	8.2280*** (0.000)	6.0084*** (0.000)	5.2585*** (0.000)	9.7445*** (0.000)	6.6276*** (0.000)	10.4832*** (0.000)	10.3404*** (0.000)
	R ²	0.8250	0.8250	0.8140	0.8670	0.8160	0.8280	0.8200	0.8300	0.8230	0.8290

Таблиця 3 (закінчення). Зв'язок між показниками десятиденних новин та інфляційними очікуваннями

Респонденти	Змінні	Інфляція		Курс валют		Комунальні послуги		Їжа		Паливо	
		Поточн.	Попер.	Поточн.	Попер.	Поточн.	Попер.	Поточн.	Попер.	Поточн.	Попер.
Суб'єкти підприємницької діяльності	π_{t-1}	0.3429*** (0.000)	0.3468*** (0.000)	0.3645*** (0.000)	0.3590*** (0.000)	0.3285*** (0.000)	0.3165*** (0.000)	0.3300*** (0.000)	0.3378*** (0.000)	0.3517*** (0.000)	0.3527*** (0.000)
	df ^m I	-0.0008 (0.999)	-0.5610 (0.409)	-0.0527 (0.940)	0.7648 (0.331)	0.9796* (0.052)	0.7777 (0.142)	1.3742 (0.104)	1.1391 (0.228)	0.7220** (0.019)	0.4026 (0.146)
	df ^m II	-0.7326 (0.355)	0.0077 (0.991)	1.1940 (0.167)	-0.7624 (0.299)	-0.0830 (0.908)	0.5752 (0.206)	0.5473 (0.558)	-0.2134 (0.865)	-0.1173 (0.682)	0.0089 (0.955)
	df ^m III	0.9513 (0.167)	0.1996 (0.723)	-0.0056 (0.995)	1.2836 (0.110)	0.2448 (0.725)	0.1232 (0.812)	-0.2146 (0.831)	-0.5889 (0.632)	-0.0325 (0.906)	0.0459 (0.853)
	C	8.0788*** (0.000)	9.5355*** (0.000)	5.9074*** (0.001)	5.8469*** (0.001)	5.7110*** (0.000)	4.7702*** (0.000)	6.5813*** (0.000)	8.1705*** (0.000)	3.9404* (0.051)	4.8362*** (0.006)
	R ²	0.6660	0.6590	0.6860	0.6890	0.7250	0.7480	0.6880	0.6640	0.6980	0.6900
Домогосподарства	π_{t-1}	0.2332*** (0.000)	0.2338*** (0.000)	0.2273*** (0.000)	0.2341*** (0.000)	0.1569*** (0.000)	0.1519*** (0.000)	0.2257*** (0.000)	0.2177*** (0.000)	0.2307*** (0.000)	0.2337*** (0.000)
	df ^m I	0.6075 (0.470)	0.3696 (0.641)	-2.1987** (0.031)	-0.6561 (0.507)	1.9786*** (0.003)	2.3171*** (0.001)	0.7161 (0.587)	1.3905 (0.280)	0.9265** (0.017)	0.6657 (0.098)
	df ^m II	-0.5300 (0.317)	-0.4296 (0.421)	0.8797 (0.287)	0.6005 (0.512)	0.6423 (0.249)	0.3836 (0.485)	2.0484* (0.088)	1.5500 (0.192)	0.3043 (0.305)	0.4797* (0.083)
	df ^m III	0.3656 (0.425)	0.3585 (0.484)	1.4200 (0.159)	0.7847 (0.463)	0.1211 (0.810)	0.2490 (0.621)	0.2029 (0.851)	0.4941 (0.682)	-0.4318 (0.120)	-0.4599 (0.136)
	C	8.7718*** (0.001)	9.1639*** (0.000)	10.3816*** (0.000)	8.9515*** (0.000)	5.5579*** (0.000)	5.2581*** (0.000)	7.0805*** (0.000)	6.7930*** (0.000)	4.4817 (0.074)	5.1578*** (0.043)
	R ²	0.6090	0.6050	0.6480	0.6140	0.7090	0.7180	0.6280	0.6270	0.6470	0.6380
Фінансові аналітики	π_{t-1}	0.1604*** (0.000)	0.1623*** (0.000)	0.1652*** (0.000)	0.1662*** (0.000)	0.1028*** (0.000)	0.0922*** (0.000)	0.1508*** (0.000)	0.1492*** (0.000)	0.1655*** (0.000)	0.1668*** (0.000)
	df ^m I	1.692*** (0.006)	1.4931** (0.011)	-1.1937 (0.134)	-0.8612 (0.263)	1.6138*** (0.001)	1.8996*** (0.000)	1.7747* (0.095)	1.4521 (0.131)	0.6686** (0.033)	0.5142* (0.100)
	df ^m II	-0.1233 (0.775)	-0.0997 (0.805)	0.5855 (0.359)	0.7450 (0.299)	0.7149* (0.069)	0.5090 (0.175)	0.4292 (0.646)	0.3017 (0.733)	0.0583 (0.796)	0.1352 (0.544)
	df ^m III	0.0259 (0.939)	0.4572 (0.218)	0.9114 (0.275)	0.4700 (0.568)	-0.0186 (0.958)	0.2341 (0.506)	0.7651 (0.354)	1.1162 (0.228)	-0.2818 (0.190)	-0.3239 (0.172)
	C	2.5993 (0.138)	1.7719 (0.325)	6.7942*** (0.000)	6.6295*** (0.000)	3.2151*** (0.000)	2.7488*** (0.000)	4.3551*** (0.000)	4.5878*** (0.000)	3.9331* (0.051)	4.7127** (0.023)
	R ²	0.6340	0.6420	0.6210	0.6110	0.7480	0.7690	0.6270	0.6230	0.6230	0.6130

Примітки: У таблиці наведено результати регресії МНК, де інфляційні очікування є залежною змінною. Перші цифри в клітинках позначають коефіцієнти, а р-значення наведено в дужках *p<0.1; **p<0.05; ***p<0.01. Поточн. означає значення за декади звітного місяця, Попер. – за попередній місяць. Індикатори I, II, III після показників частотності документа позначають кількість декад.

Таблиця 4. Зв'язок між квартальними показниками новин та інфляційними очікуваннями

Респонденти	Змінні	Без трансформацій					Змінні	1-ша різниця				
		Інфляція	Курс валют	Комунальні послуги	Їжа	Паливо		Інфляція	Курс валют	Комунальні послуги	Їжа	Паливо
Банки	π_{t-1}	0.2962*** (0.000)	0.2820*** (0.000)	0.2832*** (0.000)	0.3004*** (0.000)	0.2898*** (0.000)	$\pi_{t-1} - \pi_{t-2}$	0.0623 (0.154)	0.0903** (0.033)	0.1041** (0.015)	0.1014** (0.020)	0.0895** (0.033)
	df^m_t	-1.7001 (0.170)	1.1527 (0.467)	0.1101 (0.918)	1.3103 (0.571)	-0.0158 (0.973)	$df^m_t - df^m_{t-1}$	1.6473** (0.017)	1.5458* (0.094)	-0.4047 (0.516)	0.3991 (0.766)	0.2467 (0.368)
	df^m_{t-1}	0.7511 (0.669)	-1.7379 (0.440)	1.1679 (0.557)	-1.4879 (0.672)	-0.4129 (0.615)	$df^m_{t-1} - df^m_{t-2}$	-0.7856 (0.425)	-3.4065*** (0.010)	1.6354 (0.165)	-1.1473 (0.576)	-0.6507 (0.176)
	df^m_{t-2}	1.5584 (0.669)	-0.1744 (0.938)	-1.2997 (0.512)	-0.6431 (0.855)	-0.4109 (0.616)	$df^m_{t-2} - df^m_{t-3}$	0.0814 (0.933)	1.3132 (0.322)	-2.1069* (0.075)	-0.0197 (0.992)	0.1333 (0.782)
	df^m_{t-3}	-1.1049 (0.394)	-0.5921 (0.704)	0.8456 (0.421)	0.0465 (0.984)	0.3926 (0.399)	$df^m_{t-3} - df^m_{t-4}$	0.2624 (0.719)	0.6112 (0.508)	0.9212 (0.140)	0.7249 (0.582)	0.4089 (0.139)
	C	8.6963*** (0.000)	9.9577*** (0.000)	5.5146*** (0.000)	7.9478*** (0.000)	10.5852*** (0.000)	C	-3.6686*** (0.006)	-0.1601 (0.781)	-0.1348 (0.764)	0.0122 (0.987)	-1.0700 (0.174)
	R ²	0.8120	0.8220	0.8170	0.8100	0.8260	R ²	0.159	0.1140	0.0800	0.0590	0.1340
Суб'єкти підприємницької діяльності	π_{t-1}	0.3567*** (0.000)	0.3775*** (0.000)	0.3229*** (0.000)	0.3442*** (0.000)	0.3681*** (0.000)	$\pi_{t-1} - \pi_{t-2}$	0.1549*** (0.000)	0.1513*** (0.000)	0.1549*** (0.000)	0.1451*** (0.001)	0.1468*** (0.000)
	df^m_t	-1.3567 (0.305)	1.2189 (0.364)	0.6571 (0.428)	1.4312 (0.387)	0.3357 (0.245)	$df^m_t - df^m_{t-1}$	0.7995 (0.166)	0.8543 (0.162)	-0.3296 (0.416)	1.9479*** (0.009)	-0.0014 (0.991)
	df^m_{t-1}	0.2246 (0.914)	-0.788 (0.691)	0.7613 (0.596)	-1.2567 (0.665)	0.1356 (0.757)	$df^m_{t-1} - df^m_{t-2}$	-0.7318 (0.421)	-1.3392 (0.137)	1.3800* (0.051)	-3.1614** (0.015)	0.1989 (0.319)
	df^m_{t-2}	1.4987 (0.470)	-0.2494 (0.899)	-0.3459 (0.807)	-0.6994 (0.809)	-0.0457 (0.916)	$df^m_{t-2} - df^m_{t-3}$	0.8714 (0.335)	0.6937 (0.442)	-1.1423 (0.107)	1.8089 (0.159)	-0.3072 (0.120)
	df^m_{t-3}	0.0660 (0.961)	1.5895 (0.233)	0.4552 (0.576)	2.1718 (0.186)	0.1126 (0.676)	$df^m_{t-3} - df^m_{t-4}$	-0.8554 (0.144)	-0.0955 (0.875)	0.0518 (0.898)	-0.4610 (0.524)	0.0751 (0.541)
	C	7.4593*** (0.000)	4.5674*** (0.000)	4.6576*** (0.000)	6.6463*** (0.000)	4.0647*** (0.001)	C	-0.2782 (0.492)	-0.2835 (0.498)	0.0782 (0.798)	-0.2130 (0.475)	0.2446 (0.637)
	R ²	0.6760	0.6950	0.7460	0.6870	0.6960	R ²	0.1100	0.0880	0.1110	0.1110	0.0940

Примітки: У таблиці наведено результати регресії МНК, де інфляційні очікування є залежною змінною. Індикатор часу T частоти документа встановлено для t, t-1, t-2 і t-3 та відповідає кварталам. Перші цифри в клітинках позначають коефіцієнти, а р-значення наведено в дужках *p<0.1; **p<0.05; ***p<0.01.

Таблиця 5. Зв'язок між місячними показниками новин, визначеними ЛРД та інфляційними очікуваннями

Тема	Змінні	Без трансформацій				Змінні	1-ша різниця			
		Банки	Суб'єкти підприємницької діяльності	Домогосподарства	Фінансові аналітики		Банки	Суб'єкти підприємницької діяльності	Домогосподарства	Фінансові аналітики
Енерго-ресурси	π_{t-1}	0.2707*** (0.000)	0.3437*** (0.000)	0.2051*** (0.000)	0.1626*** (0.000)	$\pi_{t-1} - \pi_{t-2}$	0.1731*** (0.003)	0.1598*** (0.001)	0.0532 (0.291)	-0.0016 (0.970)
	df^m_t	0.4646 (0.766)	-1.0941 (0.413)	-2.7212** (0.037)	-0.3453 (0.754)	$df^m_t - df^m_{t-1}$	0.7627 (0.622)	-1.8396* (0.059)	-1.1436 (0.111)	0.3425 (0.564)
	df^m_{t-1}	-1.8021 (0.296)	2.4938* (0.072)	-1.5848 (0.196)	-0.3938 (0.708)	$df^m_{t-1} - df^m_{t-2}$	-1.0257 (0.538)	1.2913 (0.197)	1.0800 (0.116)	0.2248 (0.693)
	df^m_{t-2}	-0.9249 (0.405)	0.5825 (0.657)	-1.7524 (0.151)	-0.4304 (0.677)	$df^m_{t-2} - df^m_{t-3}$	-0.2834 (0.809)	0.2077 (0.833)	0.0709 (0.916)	0.4140 (0.455)
	df^m_{t-3}	-0.5554 (0.758)	-2.0173 (0.112)	-1.8030 (0.148)	-1.5595 (0.165)	$df^m_{t-3} - df^m_{t-4}$	-0.1006 (0.956)	-0.8172 (0.374)	0.2491 (0.249)	-1.0749* (0.070)
	C	15.9023*** (0.000)	8.9195** (0.044)	32.4739*** (0.000)	14.6931*** (0.000)	C	1.9027 (0.579)	3.4848 (0.209)	-0.7805 (0.729)	0.1552 (0.936)
	R ²	0.8460	0.6850	0.7180	0.6210	R ²	0.276	0.2150	0.0720	0.0540
Курс валют	π_{t-1}	0.2207*** (0.000)	0.2349*** (0.000)	0.0553** (0.028)	0.0443** (0.021)	$\pi_{t-1} - \pi_{t-2}$	0.1404** (0.011)	0.1296*** (0.005)	0.0376 (0.422)	-0.0148 (0.703)
	df^m_t	0.2808 (0.622)	1.9069*** (0.008)	1.2814*** (0.004)	0.7794** (0.022)	$df^m_t - df^m_{t-1}$	0.25 (0.667)	0.6776 (0.232)	0.6895** (0.020)	-0.0101 (0.968)
	df^m_{t-1}	1.0477 (0.254)	-0.6672 (0.334)	0.0629 (0.904)	0.9172** (0.032)	$df^m_{t-1} - df^m_{t-2}$	0.9799 (0.265)	-0.1744 (0.770)	-0.7571** (0.030)	0.6726** (0.031)
	df^m_{t-2}	0.7299 (0.397)	0.4674 (0.571)	0.7218 (0.165)	-0.2140 (0.608)	$df^m_{t-2} - df^m_{t-3}$	0.5198 (0.529)	0.4751 (0.509)	0.5903* (0.088)	-0.7020** (0.025)
	df^m_{t-3}	-0.4810 (0.520)	0.7308 (0.322)	1.0341** (0.022)	0.6904** (0.047)	$df^m_{t-3} - df^m_{t-4}$	-1.4949** (0.033)	-0.9657 (0.109)	-0.4956* (0.085)	0.0254 (0.917)
	C	-5.9118 (0.203)	-11.2109** (0.015)	-15.1982*** (0.000)	-10.7207*** (0.000)	C	-2.2814 (0.441)	-0.2290 (0.938)	-0.3152 (0.800)	0.0642 (0.953)
	R ²	0.8640	0.7560	0.8210	0.8090	R ²	0.397	0.2170	0.1470	0.1020

Примітки: У таблиці наведено результати регресії МНК, де інфляційні очікування є залежною змінною. Індикатор часу T частоти документа встановлено для t, t-1, t-2 і t-3. Перші цифри в клітинках позначають коефіцієнти, а р-значення наведено в дужках *p<0.1; **p<0.05; ***p<0.01.

Таблиця 6. Зв'язок між показниками новин у десятиденному вимірі, створеними ЛРД, та інфляційними очікуваннями

Респонденти	Змінні	Банки		Суб'єкти підприємницької діяльності		Домогосподарства		Фінансові аналітики	
		Поточн.	Попер.	Поточн.	Попер.	Поточн.	Попер.	Поточн.	Попер.
Енерго-ресурси	π_{t-1}	0.2788*** (0.000)	0.2715*** (0.000)	0.3554*** (0.000)	0.3647*** (0.000)	0.20600*** (0.000)	0.2151*** (0.000)	0.1588*** (0.000)	0.1605*** (0.000)
	df ^m I	-2.2081** (0.038)	-1.2279 (0.176)	0.8904 (0.394)	-0.1089 (0.914)	-2.8889*** (0.004)	-2.5453** (0.013)	-1.3360 (0.103)	-1.3255* (0.098)
	df ^m II	0.1589 (0.875)	0.6438 (0.413)	-1.9179** (0.043)	0.3250 (0.702)	-1.1732 (0.163)	-0.9812 (0.243)	-0.5387 (0.478)	0.1262 (0.858)
	df ^m III	-0.1123 (0.891)	-1.8735** (0.017)	1.0673 (0.102)	1.2336 (0.136)	-1.1909* (0.078)	-0.9189 (0.197)	0.0937 (0.882)	-0.4895 (0.384)
	C	13.8197*** (0.001)	14.9455*** (0.000)	8.3880** (0.025)	4.0943 (0.210)	25.1004*** (0.000)	22.6818*** (0.000)	12.0712*** (0.000)	11.8456*** (0.000)
	R ²	0.8380	0.8650	0.6870	0.6730	0.6890	0.6650	0.6130	0.6150
Курс валют	π_{t-1}	0.2496*** (0.000)	0.2161*** (0.000)	0.2851*** (0.000)	0.3141*** (0.000)	0.1237*** (0.000)	0.1121*** (0.000)	0.0774*** (0.000)	0.0704*** (0.000)
	df ^m I	0.8894* (0.072)	0.8441** (0.049)	1.0070** (0.014)	-0.0078 (0.987)	1.8598*** (0.000)	1.9123*** (0.000)	1.5874*** (0.000)	1.2134*** (0.000)
	df ^m II	0.8664** (0.038)	0.4398 (0.371)	-0.0442 (0.917)	0.3258 (0.463)	0.5687* (0.058)	0.7552** (0.012)	0.5167** (0.016)	0.5048** (0.022)
	df ^m III	-0.6396 (0.110)	0.3476 (0.424)	0.7195 (0.132)	0.4304 (0.328)	-0.2831 (0.373)	-0.4095 (0.184)	-0.3247 (0.132)	0.0570 (0.824)
	C	-2.2563 (0.566)	-6.3928 (0.126)	-5.2484 (0.155)	2.4127 (0.520)	-7.5191*** (0.006)	-8.4340*** (0.003)	-7.5665*** (0.000)	-7.4620*** (0.000)
	R ²	0.8490	0.8640	0.7340	0.6820	0.7520	0.7560	0.7960	0.7720

Примітки: У таблиці наведено результати регресії МНК, де інфляційні очікування є залежною змінною. Перші цифри в клітинках позначають коефіцієнти, а р-значення наведено в дужках *p<0.1; **p<0.05; ***p<0.01. Поточн. означає значення за відповідні декади звітної місяця, Попер. – за попередній місяць. Індикатори II, III після показників частоти документа позначають номер декади.

Таблиця 7. Зв'язок між квартальними показниками новин, визначеними ЛРД, та інфляційними очікуваннями

Тема	Змінні	Без трансформацій		Змінні	1-ша різниця	
		Банки	Суб'єкти підприємницької діяльності		Банки	Суб'єкти підприємницької діяльності
Енергоресурси	π_{t-1}	0.2675*** (0.000)	0.3566*** (0.000)	$\pi_{t-1} - \pi_{t-2}$	0.1846*** (0.001)	0.1466*** (0.002)
	df_t^m	-1.3148 (0.744)	1.1802 (0.661)	$df_t^m - df_{t-1}^m$	0.3972 (0.914)	-2.3983 (0.206)
	df_{t-1}^m	-2.4590 (0.609)	1.6579 (0.695)	$df_{t-1}^m - df_{t-2}^m$	-5.1964 (0.261)	4.6212 (0.137)
	df_{t-2}^m	1.2399 (0.760)	-2.7387 (0.503)	$df_{t-2}^m - df_{t-3}^m$	2.3077 (0.550)	-3.1690 (0.275)
	df_{t-3}^m	-0.5603 (0.775)	0.4886 (0.857)	$df_{t-3}^m - df_{t-4}^m$	2.1470 (0.248)	-0.0908 (0.962)
	C	16.6815*** (0.000)	6.7721 (0.148)	C	0.9729 (0.758)	3.1107 (0.271)
	R ²	0.8440	0.6620	R ²	0.3410	0.1980
Курс валют	π_{t-1}	0.2194*** (0.000)	0.2357*** (0.000)	$\pi_{t-1} - \pi_{t-2}$	-0.7182** (0.021)	0.1224*** (0.008)
	df_t^m	1.8980 (0.214)	2.8326* (0.059)	$df_t^m - df_{t-1}^m$	1.8418 (0.207)	2.6938 (0.022)
	df_{t-1}^m	-0.9758 (0.752)	-2.1270 (0.399)	$df_{t-1}^m - df_{t-2}^m$	-0.0609 (0.983)	-3.3550* (0.098)
	df_{t-2}^m	1.3292 (0.665)	0.1740 (0.941)	$df_{t-2}^m - df_{t-3}^m$	0.1379 (0.961)	1.8010 (0.341)
	df_{t-3}^m	-0.6878 (0.639)	1.4948 (0.375)	$df_{t-3}^m - df_{t-4}^m$	-1.8472 (0.156)	-1.1601 (0.388)
	C	-5.8439 (0.247)	-10.7359** (0.025)	C	0.1227 (0.807)	0.0370 (0.990)
	R ²	0.8590	0.7430	R ²	0.4320	0.2410

Примітки: У таблиці наведено результати регресії МНК, де інфляційні очікування є залежною змінною. Часовий індикатор Т частоти документа встановлюється для t, t-1, t-2 і t-3 і відповідає кварталам. Перші цифри в клітинках позначають коефіцієнти, а р-значення наведено в дужках *p<0.1; **p<0.05; ***p<0.01.

ДОДАТОК Б. РИСУНКИ

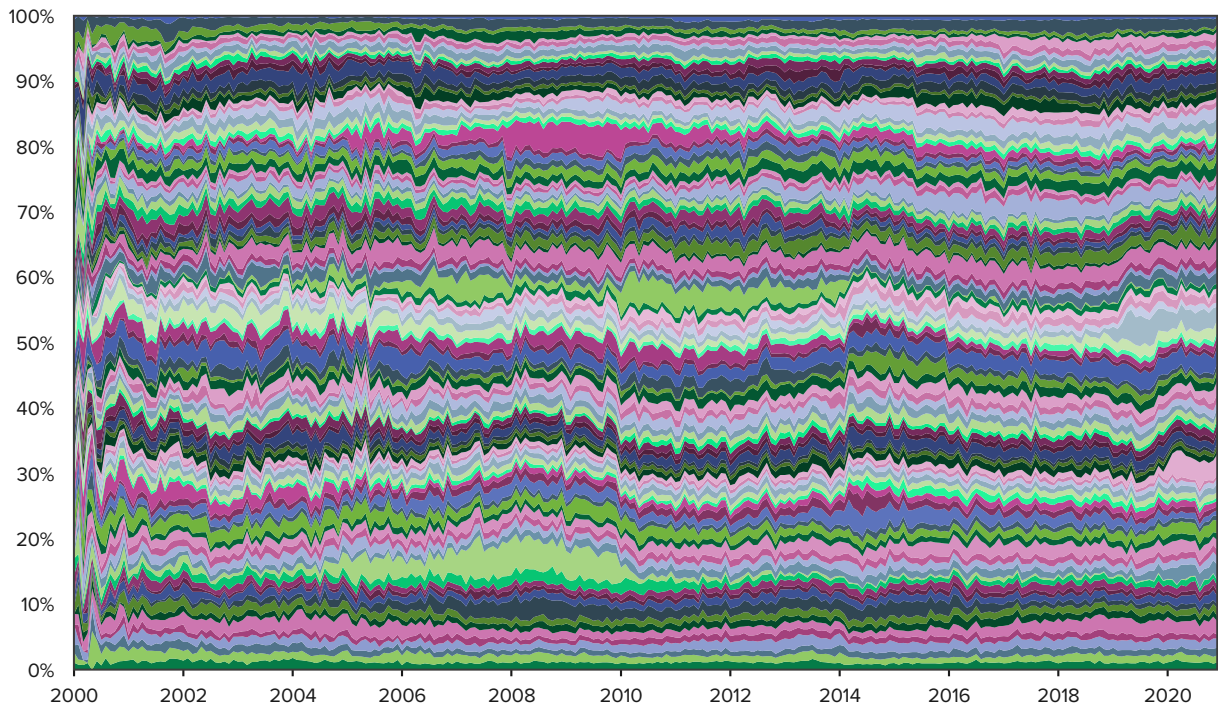


Рисунок 6. Розподіл тем, отриманих за допомогою ЛРД

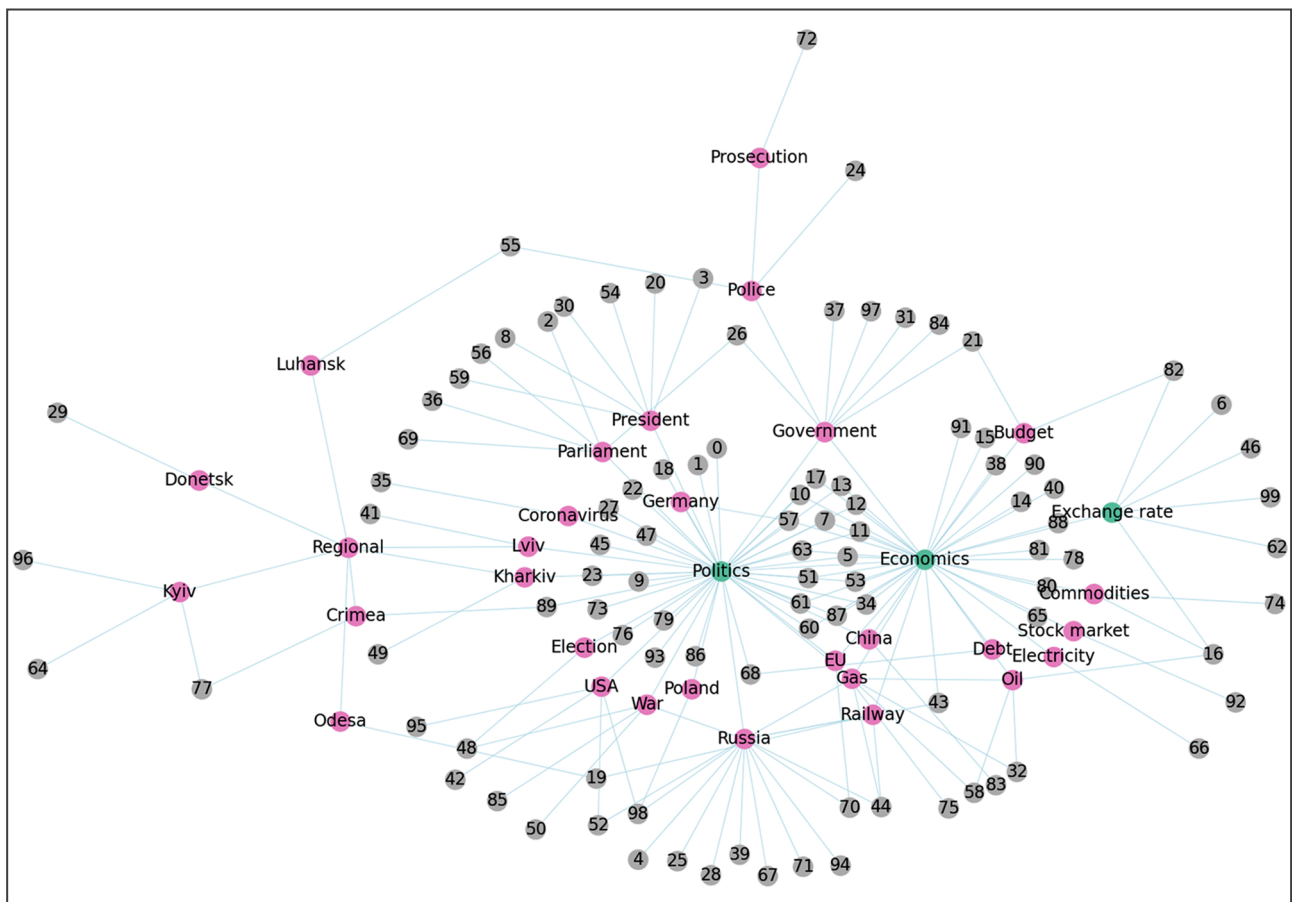


Рисунок 7. Граф тем у сукупності новин. Сірі кружечки стосуються тем, визначених ЛРД (рожеві кружечки – позначені вручну групи тем, зелені – загальні теми)

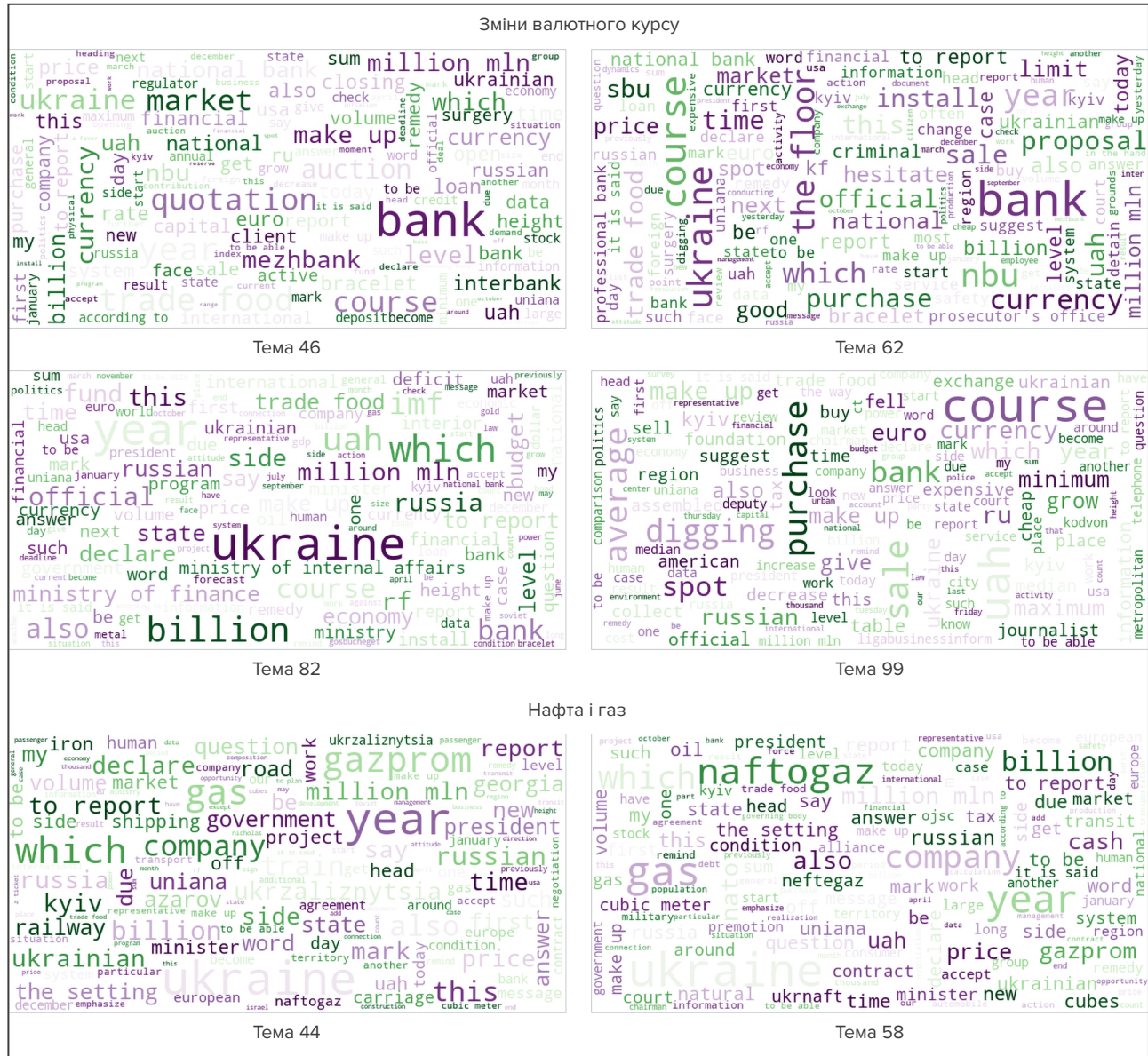


Рисунок 8. “Хмари слів” для обраних тем

ДОДАТОК В. СУКУПНІСТЬ НОВИН

Спочатку сукупність новин складалася з 2,030,000 унікальних статей. Однак після очищення та відфільтрування елементів із різними типами помилок (помилки аналізу через неправильне розміщення тегів вебсторінки, порожні сторінки, пошкоджені символи тощо) кількість статей зменшилася на 50,000 одиниць. Оскільки це лише 2.5% від загальної кількості статей, ми вважаємо таке скорочення цілком прийнятним і що воно не вплине на загальний результат.

Таблиця 8. Розмір статті в сукупності новин (після очищення)

	finance.ua	liga	ukrpravda	unian	Всього
підрахунок	389,951	620,655	339,275	634,832	1,985,143
сер.	120.6	121.5	131.0	151.5	132.5
станд.	96.8	79.6	94.2	123.9	102.1
мін.	0	2	4	3	0
25%	63	76	83	88	78
50%	99	108	115	127	113
75%	151	149	157	182	162
макс.	3,832	11,540	5,842	3,986	11,540
Коефіцієнт асиметрії розподілу	3.305	24.465	9.911	7.766	10.072
Коефіцієнт ексцесу розподілу	32.299	2,456.736	228.865	109.719	394.412

Статті в цій сукупності новин відрізняються не лише змістом, але й стилем написання, розміром, вимірною кількістю слів та іншими ознаками (табл. 8). Цілком очікувано, що різні джерела інформації мають деякі відмінності в написанні новин, що, наприклад, виявляється в розмірі статті. Найбільше в середньому статей в "УНІАН", найменше — у finance.ua. Водночас розміри статей з усіх джерел дуже схожі.

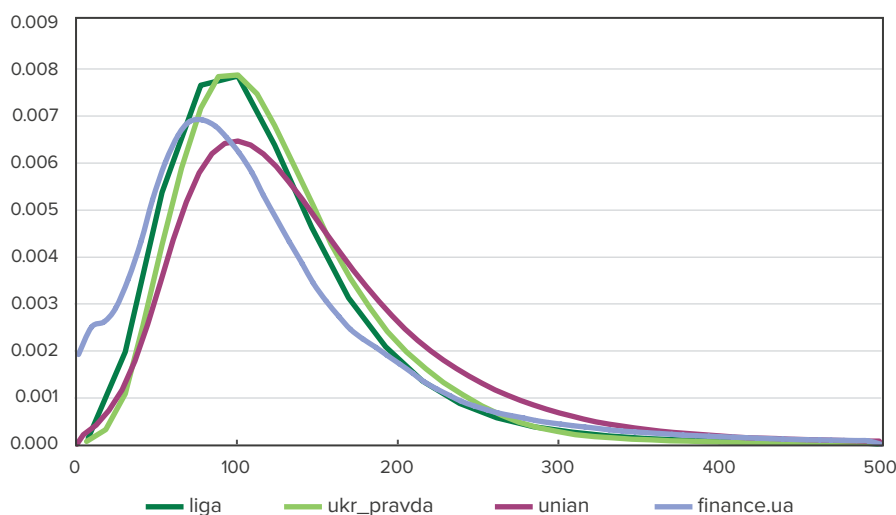


Рисунок 9. Розподіл розміру статті за джерелами

Розподіл розмірів статей для всіх джерел новин (рисунок 9) вкрай асиметричний. Усі значення асиметрії позитивні, а хвіст розподілу подовжується ближче до правого боку кривої. Статті з "Ліги" найбільш "перекошені". Водночас розподіл довжини статті є лептокуртичним, що означає, що вони високі та тонкі, отже, близькі до середнього. Наприклад, кількість статей обсягом більше 500 слів становить менше 15,000 штук, що становить лише 0.7% від усієї сукупності новин. Кількість статей дуже малого розміру також незначна (близько 2.5% від загальної кількості).

ДОДАТОК Г. РОЗРОБЛЕННЯ ДОСЛІДЖЕННЯ ІНФЛЯЦІЙНИХ ОЧІКУВАНЬ

На відміну від фінансових аналітиків, яких просять відповісти на відкрите запитання, банкам, суб'єктам підприємницької діяльності та домогосподарствам пропонується вибрати з набору інфляційних інтервалів, наприклад:

“Інфляція протягом наступних дванадцяти місяців становитиме:

- a) менше нуля (“ціни впадуть”),
- b) від 0 до X відсотків,
- c) від X до 2X відсотків,
- d) від 2X до 3X відсотків,
- e) від 3X до 4X відсотків,
- f) понад 4X відсотків.

У цьому прикладі інфляційні очікування обчислюватимуться за такою формулою:

$$E\pi = w_a \cdot \left(-\frac{X}{2}\right) + w_b \cdot \frac{X}{2} + w_c \cdot \frac{X+2X}{2} + w_d \cdot \frac{2X+3X}{2} + w_e \cdot \frac{3X+4X}{2} + w_f \cdot \left(4X + \frac{X}{2}\right), \quad (6)$$

де w – частка респондентів, які обирають відповідний інтервал. Розмір X , а також кількість інтервалів не є фіксованими та змінюються з часом відповідно до нормального розподілу відповідей. Так, у 2015 році інфляція в Україні різко прискорилося, тому максимальну планку розширили до 50%, а респондентів вибирали з 12 інтервалів. Після дезінфляції у 2020 році максимальну межу було знижено до 10%, а кількість інтервалів скорочено до восьми.

З січня 2018 року опитування домогосподарств також включають питання про сприйняття інфляції. Раз на рік споживачів просять відповісти на одне відкрите запитання про очікувану інфляцію за наступні 12 місяців. Крім того, раз на квартал домогосподарствам пропонується вибрати відповіді на інтервальне питання. Формулювання цього питання аналогічне питанню інфляційних очікувань.