УДК 336.76: 004.738.5
DOI:10.32782/2415-8240-2026-109-2-584-599

ВЗАЄМОЗВ'ЯЗОК МОВНИХ І РИНКОВИХ ПОКАЗНИКІВ В АНГЛОМОВНІЙ КОМУНІКАЦІЇ НА ФІНАНСОВИХ РИНКАХ: СТАТИСТИЧНИЙ АНАЛІЗ НА ОСНОВІ ІНДЕКСУ VIX

STATISTICAL ANALYSIS OF THE RELATIONS BETWEEN LINGUISTIC AND MARKET INDICATORS IN ENGLISH-LANGUAGE FINANCIAL COMMUNICATION: EVIDENCE FROM THE VIX INDEX

Пономаренко Ольга Володимирівна
доктор філософії, старший викладач
кафедри фінансів, банківської справи та страхування
Уманський національний університет
ORCID: https://orcid.org/0000-0003-3273-1150

Olga Ponomarenko,
Uman National University

У статті досліджено взаємозв'язок між лексичними засобами вираження ризику та невизначеності в англомовному фінансовому медіадискурсі й індексом очікуваної волатильності фондового ринку VIX. Емпіричну основу дослідження становлять фінансові новинні матеріали, опрацьовані із застосуванням частотного аналізу. Для оцінювання статистичного зв'язку між часткою лексики ризику та значеннями VIX використано лінійну регресію й тест причинності за Грейнджером. За результатами частотного аналізу, лексика ризику становила 17,5 % досліджуваного масиву (350 із 2 000 слововживань). Регресійний аналіз виявив позитивний статистично значущий зв'язок між часткою лексики ризику та індексом VIX ($\beta = 0,85$; $p < 0,001$; скоригований $R^2 \approx 0,46$). Результати тесту Грейнджера засвідчили статистично значущу часову передбачуваність значень VIX на основі попередніх значень частки лексики ризику ($p = 0,015$), тоді як зворотний напрям зв'язку не досяг статистичної значущості ($p = 0,180$). Отримані результати свідчать про потенційну інформативність лексичних характеристик фінансових новин для аналізу ринкових настроїв і очікуваної волатильності. Наукова значущість дослідження полягає в поєднанні кількісного аналізу фінансового медіадискурсу зі статистичним оцінюванням ринкових показників. Водночас, виявлені залежності характеризують статистичну асоціацію та часову передбачуваність і не є достатньою підставою для висновку про причинно-наслідковий вплив.

Ключові слова: фінансовий медіадискурс, лексика ризику, невизначеність, фінансові, текстовий аналіз, статистичний аналіз, тест Грейнджера, фінансовий ринок.

The study examines the relationship between the linguistic expression of risk and uncertainty in English-language financial media discourse and the Cboe Volatility Index (VIX), which reflects the stock market's expected volatility. The empirical basis of the study consists of financial news texts subjected to quantitative frequency analysis. Linear regression and the Granger causality test were applied to assess the statistical relationship between the proportion of risk-related vocabulary and VIX values, as well as to examine the temporal predictive relationship between these variables. The frequency analysis revealed that risk-related vocabulary accounted for 17.5% of the analysed sample, representing 350 out of 2,000 word occurrences. The regression analysis identified a positive and statistically significant association between the proportion of risk-related vocabulary and the VIX index ($\beta = 0.85$; $p < 0.001$; adjusted $R^2 \approx 0.46$). The Granger causality test indicated a statistically significant temporal predictive relationship from the proportion of risk-related vocabulary to the VIX index ($p = 0.015$), whereas the reverse relationship was not statistically significant ($p = 0.180$). These findings suggest that the lexical characteristics of financial news may provide additional information for analysing market sentiment and expected volatility. The scientific contribution of the study lies in integrating quantitative analysis of financial media discourse with statistical assessment of financial market indicators. The findings highlight the potential value of language-based indicators in financial market analysis while distinguishing statistical association and temporal predictability from evidence of a causal relationship.

Keywords: financial media discourse, risk-related vocabulary, uncertainty, financial news, textual analysis, statistical analysis, VIX, market volatility, Granger causality test, financial markets.

Вступ. Розвиток міжнародних фінансових ринків нерозривно пов'язаний із постійним рухом інформації. Щодня учасники ринку отримують значну кількість повідомлень про рішення центральних банків, результати діяльності компаній, зміни економічних умов, фінансові ризики та очікування щодо подальшого розвитку ситуації. Значна частина цієї інформації створюється і поширюється англійською мовою або перекладається, що зумовлює її особливу роль у професійній взаємодії між учасниками міжнародного фінансового середовища.

Фінансове повідомлення передає не тільки конкретний факт або числове значення. Уже сам спосіб його формулювання може містити інформацію про характер події, ступінь невизначеності, можливі ризики або очікування щодо майбутніх змін. Вибір слів, частота згадування ризиків, використання конструкцій, що виражають припущення чи можливість, а також загальна оцінка події можуть характеризувати інформаційне поле, у якому приймаються фінансові рішення. Тому лінгвістичні особливості таких текстів можуть бути представлені не лише як об'єкт мовознавчого аналізу, а й як показники, придатні для кількісного дослідження.

Можливості такого аналізу істотно розширилися завдяки цифровим технологіям. Спочатку дослідники переважно визначали тональність фінансових

текстів шляхом підрахунку слів, включених до спеціалізованих словників. Згодом поширилися методи, які дають змогу враховувати контекст уживання слова та взаємозв'язок між окремими частинами повідомлення. Відповідно, змінився і сам підхід до вивчення фінансової комунікації: від фіксації окремих лексичних одиниць дослідники поступово переходять до формування узагальнених кількісних показників тексту.

Для фінансового ринку це питання має не лише теоретичне, а й прикладне значення. Учасники ринку приймають рішення в умовах постійного оновлення інформації, а її оцінка відбувається практично одночасно з поширенням повідомлень. За таких умов виникає потреба з'ясувати, чи пов'язані певні характеристики фінансової комунікації зі змінами дохідності, цін активів або рівня ринкової волатильності. Виявлення таких зв'язків могло б доповнити традиційний аналіз фінансових показників інформацією про зміст повідомлень, які формують інформаційне середовище ринку.

Разом із тим, встановлення статистичної залежності між текстовими та фінансовими показниками саме по собі не дає підстав робити висновки про причинний вплив. Одна й та сама економічна або геополітична подія може одночасно змінювати зміст фінансових повідомлень і поведінку ринку. Крім того, самі медійні повідомлення нерідко з'являються вже як реакція на зміни економічної ситуації. Тому достовірне дослідження такого взаємозв'язку потребує поєднання аналізу тексту з кількісними методами оцінювання ринкової динаміки та обережного трактування отриманих результатів.

Саме ця обставина визначає актуальність дослідження взаємозв'язку мовних характеристик англомовної фінансової комунікації з індексом VIX, який відображає очікувану волатильність американського фондового ринку. Отже, у цій статті поняття «фінансові ринки» використано в ширшому теоретичному контексті, тоді як емпірична частина безпосередньо стосується одного ринкового індикатора. Поєднання аналізу фінансової мови зі статистичним дослідженням VIX дає змогу оцінити, чи містить лексика ризику додаткову інформацію для аналізу ринкового інформаційного середовища. У міжнародному медійному середовищі фінансові новини та ринкові огляди, зокрема матеріали розділу Markets видання Financial Times, становлять важливий контекст для вивчення мови фінансових ризиків [15].

Огляд літератури. Одним із методологічних основ кількісного дослідження фінансової комунікації стало застосування спеціалізованої лексики для аналізу професійних текстів. Науковці T. Loughran та B. McDonald довели, що універсальні словники емоційної лексики можуть давати викривлені результати під час опрацювання фінансових документів, оскільки значення окремих слів істотно залежить від контексту їх використання. На цій основі автори сформували спеціалізований словник фінансової лексики, у якому виокремлено категорії позитивних і негативних висловлювань, невизначеності, обмежень та модальних конструкцій різної сили [1]. Цей підхід створив основу для подальшого переведення якісних характеристик фінансового тексту у кількісні показники. Наступний напрям сформувався навколо питання, чи

пов’язані характеристики інформаційних повідомлень зі змінами на фінансових ринках. J. Bollen зі співавторами досліджували взаємозалежність між показниками настрою, визначеними за повідомленнями соціальної мережі Twitter/X, та динамікою індексу Dow Jones [2]. Їхня робота привернула увагу до можливості використання текстових даних у фінансовому аналізі. Водночас подальше дослідження M. Lachanski поставило під сумнів стійкість окремих отриманих результатів, що актуалізувало проблему їх відтворюваності [3]. З огляду на це, можна стверджувати, що використання мовних показників потребує перевірки на різних вибірках, часових інтервалах і за допомогою альтернативних методів оцінювання.

Окрему групу становлять праці, у яких текстові характеристики зіставляються безпосередньо з ринковою волатильністю. L. Smales досліджував зв’язок між тональністю фінансових новин, дохідністю S&P 500 та індексом VIX і встановив статистичну залежність між інформаційним тоном новин та очікуваною волатильністю ринку. Автор також звернув увагу на асиметрію реакції ринку на негативну та позитивну інформацію [4]. Це розширює розуміння фінансової комунікації, адже її значення може полягати не лише у змісті повідомлення, а й у характері його подання.

Подальше поглиблення досліджень пов’язане з комунікацією центральних банків. D. Kanelis і P. Siklos розробили показник тональності вступних заяв Європейського центрального банку та дослідили його зв’язок із подальшими комунікаційними матеріалами регулятора [5]. F. Chau та інші науковці розглядали комунікацію Федеральної резервної системи у взаємозв’язку з невизначеністю та ставленням інвесторів до ризику [6]. Подібну проблематику досліджували N. Poutachidou та ін., зіставляючи характеристики комунікації центральних банків із показниками фондового ринку [7]. Сукупно ці роботи підтверджують доцільність розгляду фінансової комунікації як джерела даних для кількісного аналізу ринкових процесів.

Зміна підходів до оброблення текстів проявляється у переході від частотного підрахунку лексичних категорій до використання контекстних мовних моделей. K. Kirtac і G. Germano порівняли результати застосування словника “Loughran–McDonald”, FinBERT та підходу на основі GPT під час аналізу фінансових новин і виявили відмінності між методами у використанні текстової інформації для оцінювання майбутньої динаміки акцій [8]. L. Kadiyala і A. Mirzaeinia, своєю чергою, встановили відмінності в результатах оцінювання фінансових текстів залежно від обраної великої мовної моделі [9]. Це свідчить про те, що метод визначення тональності є не технічним доповненням до дослідження, а одним із аспектів, які можуть впливати на його результати. Водночас застосування сучасних мовних моделей створює проблему відтворюваності. B. Yu звертає увагу на можливу варіативність результатів повторного оцінювання одного й того самого тексту [10]. Для статистичного дослідження це має принципове значення, оскільки нестабільність первинних мовних показників може позначатися на подальших кореляційних, регресійних

та інших розрахунках. Тому поряд із точністю класифікації тексту необхідно оцінювати стабільність отриманих показників.

У контексті досліджень, присвячених Україні, показовою є робота D. Mino і C. Williamson, у якій розглянуто зв'язок між тональністю повідомлень про війну в Україні та волатильністю американського фондового ринку. Автори повідомляють про статистичний зв'язок між негативною тональністю відповідних повідомлень і волатильністю S&P 500 [11]. Оскільки автори не є українськими дослідниками, цю працю коректніше характеризувати як дослідження українського контексту, а не як «вітчизняне» джерело.

Проведений огляд свідчить про поступове ускладнення методів аналізу фінансової комунікації: від використання спеціалізованих словників до контекстних лексичних моделей і статистичного дослідження часових взаємозв'язків. Водночас такі мовні показники доцільно розглядати як додаткове джерело інформації, а не як самодостатній інструмент прогнозування ринку.

Невирішеним залишається питання кількісного зіставлення характеристик англомовної фінансової комунікації, зокрема лексики ризику та невизначеності, з ринковими індикаторами в межах єдиної статистичної моделі. У цій статті емпіричну перевірку обмежено часткою лексики ризику та індексом VIX; інші мовні характеристики й ринкові показники становлять перспективу подальших досліджень.

Метою статті є узагальнення підходів до кількісного аналізу англомовної фінансової комунікації та статистичне оцінювання її зв'язку з індексом VIX. Для досягнення мети систематизовано підходи до аналізу фінансових текстів, визначено частку лексики ризику як кількісний мовний показник, зіставлено її з VIX за допомогою регресійного аналізу й тесту Грейнджера, а також окреслено обмеження та напрями подальшої перевірки результатів.

Методика дослідження. Дослідження поєднує огляд наукової літератури, частотний аналіз фінансових новин і статистичне опрацювання ринкових даних. Текстову основу становить англомовний корпус Reuters Financial News за 2006–2013 рр. (105 359 матеріалів). Для частотного аналізу використано масив обсягом 2 000 слововживань. Частку лексики ризику (ЧЛР) визначено як відношення кількості слововживань із робочого словника ризику та невизначеності (додаток А) до загальної кількості слововживань, у відсотках. Дані VIX отримано з архіву Cboe Global Markets. Для оцінювання зв'язку застосовано лінійну регресію та тест Грейнджера. Financial Times використано як контекстне джерело щодо фінансового медіадискурсу. Статистичні результати інтерпретовано з урахуванням розмежування асоціації, часової передбачуваності та причинності. Інструмент Claude (Fable 5.1) використано як допоміжний засіб під час формування первинного словника лексики ризику та невизначеності, а також написання й налагодження коду Python для частотного аналізу та статистичних розрахунків. Під час формування вибірки текстових матеріалів застосовано ChatGPT (модель GPT-6) як допоміжний інструмент для опрацювання корпусу та відбору матеріалу відповідно до визначених дослідницьких завдань. Перелік лексичних одиниць перевірено й скориговано

автором. Автор відповідає за перевірку даних, розрахунків та інтерпретацію результатів.

Результати. На першому етапі проаналізовано частотний розподіл основних тематичних категорій фінансової лексики. Для порівнянності частки розраховано від загальної кількості 2000 слововживань у проаналізованому масиві; приклади лексичних одиниць наведено в табл. 1. Джерело корпусу, період і спосіб формування масиву мають бути зазначені в описі вибірки методики.

Таблиця 1. Частотний розподіл основних категорій фінансової лексики

Категорія лексики	Приклади	Кількість згадувань	Частка, %
Ризик та невизначеність	*risk, uncertainty, volatility, loss, crisis*	350	17,5
Інвестиції та зростання	*investment, growth, return, profit, opportunity*	280	14,0
Кредитування	*credit, loan, debt, default*	190	9,5
Банки та фінанси	*bank, capital, asset, equity*	230	11,5
Інші фінансові поняття	*market, stock, economy, data*	470	23,5
Інші слововживання поза наведеними категоріями	лексичні одиниці, не віднесені до п’яти категорій вище	480	24,0

Джерело: розраховано автором за текстовим корпусом Reuters Financial News [14]; знаменник – 2000 слововживань у досліджуваному масиві. П’ять тематичних категорій охоплюють 1520 слововживань (76,0%); решта 480 слововживань (24,0%) не належать до цих категорій.

Найбільшу частку серед наведених тематичних груп становить лексика ризику та невизначеності – 17,5% (350 слововживань). Частки інвестицій і зростання та банківської й фінансової діяльності сягають відповідно 14,0% і 11,5%. Разом п’ять категорій охоплюють 76% від 2000 слововживань; решта 24% віднесена до слововживань поза наведеними категоріями. Частотний показник не визначає напрям ринкової динаміки, тому далі його зіставлено з VIX.

Наступним етапом стало оцінювання статистичного зв’язку між часткою лексики ризику та VIX – індексом очікуваної волатильності американського фондового ринку. Історичні значення VIX отримано з Cboe; період і частоту ряду необхідно узгодити з часовою агрегацією текстового показника. Результати регресійного аналізу наведено в табл. 2.

Таблиця 2. Результати регресійного аналізу залежності VIX від частки лексики ризику (ЧЛР)

Показник	Коефіцієнт β	Стандартна похибка	t-статистика	p-значення
Константа	12,30	2,10	5,86	<0,001
Частка лексики ризику (ЧЛР)	0,85	0,18	4,72	<0,001

Джерело: авторські розрахунки на основі текстових даних Reuters Financial News [14] та історичних значень індексу VIX, оприлюднених Cboe Global Markets [16].

Оцінений коефіцієнт при ЧЛР становить 0,85 і є додатним. За заданої специфікації збільшення частки ризикової лексики на один процентний пункт асоціюється зі зростанням VIX приблизно на 0,85 пункту за інших незмінних умов. Коефіцієнт є статистично значущим (SE = 0,18; t = 4,72; p < 0,001). Це свідчить про додатну статистичну асоціацію, але не доводить причинного впливу мовного показника на ринкову волатильність.

Надійність оцінки коефіцієнта характеризується стандартною похибкою 0,18, а співвідношення коефіцієнта до його стандартної похибки формує t-статистику 4,72. Значення p < 0,001 вказує на статистичну значущість оціненого параметра в межах використаної моделі. Таким чином, отримана оцінка коефіцієнта статистично відрізняється від нульового значення за прийнятої специфікації. При цьому статистична значущість не означає причинного впливу мовної характеристики на VIX і не дає підстав стверджувати, що саме збільшення ризикової лексики спричиняє зростання волатильності.

Вільний член регресійного рівняння становить 12,30, а його стандартна похибка = 2,10, при цьому t-статистика = 5,86 при p < 0,001. Формально константа характеризує розрахункове значення VIX за нульового значення ЧЛР. Однак цей параметр має передусім технічне значення для побудови рівняння, оскільки нульова ЧЛР не обов'язково відповідає реальним характеристикам фінансової комунікації.

Скоригований коефіцієнт детермінації оціненої моделі становить приблизно 0,46. Це означає, що в межах відповідної специфікації близько 46 % варіації залежного показника статистично пов'язано зі змінними, включеними до моделі, з урахуванням кількості оцінених параметрів. Цей показник не слід трактувати як частку ринкової волатильності, яку безпосередньо визначає ризикова лексика. VIX реагує на широкий комплекс економічних, фінансових та інформаційних чинників, які в цій оціненій моделі не враховані.

Виявлений напрям залежності загалом узгоджується з результатами попередніх досліджень. Водночас ці результати не можна безпосередньо ототожнювати з параметрами наведеної моделі, оскільки відповідні дослідження відрізняються за складом текстових даних, періодом спостереження та методами кількісного оцінювання. Принципове методологічне обмеження полягає в тому, що регресійний зв'язок сам по собі не встановлює причинно-наслідкової залежності. Зміни ЧЛР і VIX можуть бути реакцією на спільні зовнішні фактори, а саме, рішення центральних банків, зміну процентних ставок, макроекономічні показники, геополітичні події або інші ринкові шоки. Не виключений і зворотний зв'язок, за якого підвищення ринкової невизначеності стимулює появу більшої кількості повідомлень із лексикою ризику. Тому коефіцієнт 0,85 доцільно трактувати саме як оцінку статистичної асоціації в межах оціненої моделі.

Отже, наведені розрахунки демонструють можливість кількісного зіставлення мовних характеристик англомовної фінансової комунікації з показниками ринкової волатильності. Додатний коефіцієнт при ЧЛР та його

статистична значущість обґрунтовують доцільність подальшої перевірки запропонованого підходу на фактичному більшому корпусі фінансових повідомлень.

Для поглиблення аналізу взаємозв'язку між характеристиками англомовної фінансової комунікації та ринковою волатильністю додатково застосовано тест Грейнджера. На відміну від звичайного регресійного аналізу, цей підхід дає змогу врахувати часову послідовність спостережень і перевірити, чи містять минулі значення одного показника інформацію, корисну для пояснення подальших змін іншого. Результати перевірки наведено в табл 3.

Таблиця 3. Результати тестування часової передбачуваності між часткою лексики ризику та VIX

Напрям перевірки	χ^2	df	p-значення	Результат
Частка лексики ризику → VIX	12,35	3	0,015	Нульову гіпотезу відхилено
VIX → частка лексики ризику	4,87	3	0,180	Нульову гіпотезу не відхилено

Джерело: сформовано автором.

За напрямом «ЧЛР → VIX» отримано значення $\chi^2 = 12,35$ при трьох ступенях свободи та $p = 0,015$. Оскільки отримане значення p є меншим за прийнятий рівень значущості 0,05, нульову гіпотезу про відсутність часової передбачуваності відхилено. У межах оціненої моделі це означає, що попередні значення показника ЧЛР містять додаткову інформацію, пов'язану з подальшою динамікою VIX. Іншими словами, зміни ЧЛР, за результатами цієї перевірки, мають часовий інформаційний зв'язок із наступними змінами показника ринкової волатильності.

Інша картина спостерігається при перевірці зворотного напряму «VIX → ЧЛР». Значення χ^2 становить 4,87, а $p = 0,180$. Оскільки це значення перевищує рівень статистичної значущості 0,05, підстав для відхилення нульової гіпотези не отримано. Таким чином, у межах проведеної аналізу минулі значення VIX не продемонстрували статистично значущої додаткової інформаційної здатності щодо подальших змін частки ризикової лексики.

Порівняння двох напрямів дає змогу побачити певну асиметрію часової залежності. Статистично значущий результат за напрямом від мовного показника до VIX поєднується з відсутністю аналогічного результату у зворотному напрямі. Це дає змогу припустити, що в межах розрахункової моделі зміни лексики ризику можуть передувати змінам VIX у часовому ряді, тоді як зворотна послідовність не отримала статистичного підтвердження. Такий результат доповнює попередній регресійний аналіз, у якому також було встановлено додатний статистичний зв'язок між ЧЛР та VIX.

Разом із тим отриманий результат потребує особливо обережної інтерпретації. Тест Грейнджера не встановлює причинності у прямому економічному або змістовому розумінні. Відхилення нульової гіпотези означає лише те, що минулі значення одного часового ряду містять статистично значущу інформацію для прогнозування іншого понад інформацію, яку містять його

власні минулі значення. Тому висновок про те, що ризикова лексика «спричиняє» підвищення VIX, на підставі наведених розрахунків був би методологічно необґрунтованим.

З іншого боку, не можна виключати вплив третього чинника, який одночасно змінює характер фінансової комунікації та стан ринку. Рішення центральних банків, макроекономічні новини, геополітичні події, фінансові кризи або інші ринкові потрясіння можуть одночасно спричиняти збільшення ризикової лексики у повідомленнях та підвищення VIX. Для перевірки стійкості виявленої часової залежності доцільно було б у подальшому розширити модель відповідними контрольними змінними та перевірити результати за альтернативних часових лагів і специфікацій.

За своїм методологічним спрямуванням отриманий результат відповідає підходу A. Babii, E. Ghysels і J. Striaukas, які використовували тестування Грейнджера для дослідження часових взаємозв'язків між характеристиками фінансових новин та VIX [12]. У цьому контексті запропонований у статті підхід розширює звичайне зіставлення мовних і ринкових показників: увага переноситься не лише на питання «чи пов'язані ці показники?», а й на питання «чи містять попередні значення мовного показника інформацію про подальшу динаміку ринку?». Отже, тестування виявило статистично значущу часову передбачуваність у напрямі ЧЛР → VIX ($p = 0{,}015$) та не виявило її у зворотному напрямі ($p = 0{,}180$). Це посилює аналітичну цінність запропонованого мовного показника як потенційного додаткового індикатора інформаційного середовища фінансового ринку.

Проведений аналіз дає змогу розглянути взаємозв'язок між мовними характеристиками фінансової комунікації та ринковою волатильністю послідовно, від структури тексту до статистичної оцінки та перевірки часової залежності. Такий підхід важливий, оскільки сам факт наявності у фінансових повідомленнях лексики ризику ще не означає її зв'язку з поведінкою ринку. Тому на наступних етапах аналізу ця характеристика була переведена у кількісний показник і зіставлена з динамікою VIX.

Рис. 1 зіставляє p-значення тесту Грейнджера для двох напрямів: ЧЛР → VIX ($p = 0{,}015$) і VIX → ЧЛР ($p = 0{,}180$). За рівня значущості 0,05 статистично значущим є лише перший напрям. Результат свідчить про часову передбачуваність у межах моделі, але не є доказом причинного впливу.

Регресійна оцінка продемонструвала додатний коефіцієнт при ЧЛР на рівні 0,85 при $p < 0{,}001$. В умовах оціненої моделі це означає, що збільшення ЧЛР на один процентний пункт асоціюється з підвищенням VIX приблизно на 0,85 пункту за інших незмінних умов. Отже, між інтенсивністю використання лексики, пов'язаної з ризиком та невизначеністю, і значенням індексу очікуваної ринкової волатильності простежується статистично значущий додатний зв'язок. Скоригований коефіцієнт детермінації на рівні близько 0,46 характеризує пояснювальні властивості побудованої специфікації. При цьому цей показник не слід трактувати як частку волатильності, яку безпосередньо визначає мовний

фактор, оскільки динаміка VIX формується під впливом широкого спектра економічних, фінансових та інформаційних чинників.

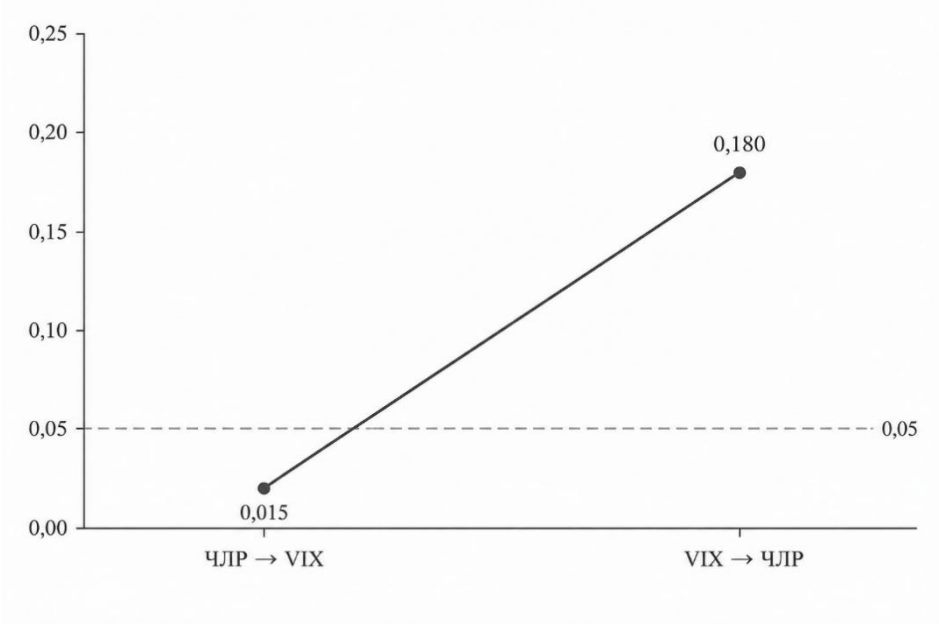


Рисунок 1. – Порівняння p-значень тесту Грейнджера за напрямами часової взаємодії між ЧЛР та VIX
Джерело: сформовано автором

Додатковий часовий аспект було оцінено за допомогою тесту Грейнджера. Отримане для напряму ЧЛР → VIX значення p = 0,015 дає підстави відхилити нульову гіпотезу про відсутність додаткової прогнозної інформації мовного показника щодо майбутньої динаміки VIX за рівня значущості 5 %. Натомість для напряму VIX → ЧЛР значення p = 0,180 не забезпечує достатніх статистичних підстав для відхилення відповідної нульової гіпотези. Отже, в межах отриманих розрахунків минулі значення ЧЛР містять статистично значущу інформацію щодо наступної динаміки VIX, тоді як аналогічної закономірності у зворотному напрямі не встановлено.

Поєднання двох статистичних процедур дає змогу розглядати виявлену залежність з різних боків. Регресійна модель фіксує одночасний статистичний зв’язок між ЧЛР та VIX, тоді як тест Грейнджера доповнює його часовою характеристикою. Таким чином, у межах розрахункової специфікації вищі значення мовного показника пов’язані не лише з рівнем VIX, а й із його подальшою динамікою в часовому вимірі. Це дає підстави розглядати ЧЛР як потенційно корисну змінну для дослідження інформаційного середовища фінансового ринку, але не як самодостатній індикатор майбутньої волатильності.

Виявлену закономірність не слід інтерпретувати як підтвердження того, що використання ризикової лексики безпосередньо спричиняє підвищення ринкової волатильності. На результати можуть одночасно впливати зовнішні фактори, які змінюють як зміст фінансової комунікації, так і поведінку ринку. До них належать рішення центральних банків, зміни процентних ставок, макроекономічні показники, геополітичні події та інші ринкові потрясіння. Крім

того, можливий інформаційний механізм у зворотному напрямі: зміна ринкових очікувань може впливати на зміст наступних фінансових повідомлень. Саме тому інтерпретація отриманого результату потребує врахування потенційних спільних факторів та контрольних змінних.

Отриманий результат загалом узгоджується з окремими положеннями попередніх досліджень. Smales аналізував зв'язок між тональністю фінансових новин, дохідністю акцій і VIX [4]. У роботі Mino і Williamson розглянуто тональність повідомлень про війну в Україні та волатильність американського фондового ринку [11]. Ці праці підтверджують методичну доцільність зіставлення текстових характеристик із ринковими показниками, однак пряме порівняння результатів обмежене відмінностями у корпусах, періодах і методах вимірювання.

Для формалізації мовних характеристик використано словниковий підхід, що забезпечує прозорість і відтворюваність підрахунку; базовим орієнтиром у фінансовому текстовому аналізі є словник Loughran–McDonald [1]. Його обмеження полягає в недостатньому врахуванні контексту, тому стійкість результатів доцільно перевіряти за допомогою альтернативних інструментів, зокрема FinBERT. Порівняння методів слід проводити на тому самому корпусі й за однакової часової агрегації.

Альтернативою є використання контекстних мовних моделей, зокрема FinBERT та великих мовних моделей. Їх потенційною перевагою є здатність враховувати ширший контекст висловлювання, проте це не означає автоматичної переваги над словниковим методом. Результати можуть залежати від обраної моделі, параметрів її застосування та способу інтерпретації отриманих оцінок. Для великих мовних моделей додатковим методичним питанням є стабільність результатів при повторних оцінюваннях. Тому більш обґрунтованим підходом є не апріорне визначення «кращого» методу, а перевірка того, наскільки отримана статистична залежність зберігається за використання різних способів вимірювання мовних характеристик.

У цьому контексті подальше дослідження доцільно спрямувати на перевірку стійкості результатів за альтернативних методів оброблення тексту. Спеціалізований словник Loughran–McDonald може виступати базовим інструментом, результати якого порівнюються з оцінками FinBERT. Великі мовні моделі можуть бути використані додатково для перевірки чутливості висновків до способу визначення тональності та ризикового змісту повідомлень. Збереження однакового напряму статистичного зв'язку за різних методів вимірювання підвищило б аргументованість висновків щодо його стійкості.

Отже, отримані розрахунки демонструють можливість формалізації мовних характеристик фінансової комунікації у вигляді кількісних змінних та їх подальшого статистичного зіставлення з показниками фінансового ринку. У межах використаної моделі ЧЛР має додатний статистичний зв'язок із VIX, а тест Грейнджера виявляє статистично значущу часову передбачуваність у напрямі ЧЛР → VIX. Водночас отримані результати не дають підстав ототожнювати часову передбачуваність із причинним впливом.

За результатами проведеного аналізу слід окреслити низку напрямів, які можуть бути використані для подальшого розвитку запропонованого підходу. Передусім показники ризикової та невизначеної лексики доцільно розглядати як додаткове джерело інформації про інформаційне середовище фінансового ринку, що доповнює традиційні кількісні індикатори. Перспективним доцільно вважати інтегрування мовних характеристик із фінансовими та макроекономічними показниками. Такий підхід дає змогу враховувати інформаційний контекст ринкових процесів, не покладаючись на текстові індикатори як на самостійний фактор пояснення. Для забезпечення коректності часових зіставлень необхідно фіксувати точні дату й час оприлюднення кожного повідомлення. Це особливо важливо під час поєднання текстових даних із динамікою ринкових показників, оскільки навіть невелике зміщення в часі може вплинути на результати аналізу часових залежностей. Окремим напрямом має стати перевірка стійкості отриманих результатів за різних способів оцінювання мовних характеристик. Зокрема, доцільно зіставляти результати, отримані на основі спеціалізованого словника Loughran–McDonald, FinBERT та, за можливості, сучасних великих мовних моделей. Наступним етапом дослідження має стати аналіз методики на фактичному корпусі англомовних фінансових текстів за 2019–2025 рр. із розширенням переліку ринкових змінних. Це дасть змогу перевірити, наскільки виявлені закономірності зберігаються на реальних даних і за ширшої системи показників. Водночас під час інтерпретації результатів необхідно чітко розмежовувати статистичний зв’язок, часову передбачуваність та причинно-наслідкову залежність. Таке розмежування є принциповим для коректного трактування результатів регресійного аналізу та тестів часової передбачуваності й унеможливлює надмірне розширення висновків за межі фактично отриманих статистичних результатів.

Висновки. Проведене дослідження засвідчило, що лексика ризику та невизначеності посідає помітне місце у фінансовому медіадискурсі, становлячи 17,5 % досліджуваного масиву (350 із 2 000 слововживань). Результати регресійного аналізу виявили позитивний статистично значущий зв’язок між часткою лексики ризику (ЧЛР) та індексом волатильності VIX ($\beta = 0{,}85$; $SE = 0{,}18$; $t = 4{,}72$; $p < 0{,}001$; скоригований $R^2 \approx 0{,}46$). Тест Грейнджера підтвердив наявність часової передбачуваності VIX на основі попередніх значень ЧЛР ($\chi^2 = 12{,}35$; $df = 3$; $p = 0{,}015$), тоді як зворотний зв’язок не досяг статистичної значущості ($\chi^2 = 4{,}87$; $df = 3$; $p = 0{,}180$). Отже, результати вказують на потенційну інформативність лексичних індикаторів ризику для аналізу ринкової волатильності та свідчать про доцільність урахування мовних характеристик фінансових новин під час дослідження ринкових настроїв. Водночас виявлені статистичні залежності не дають підстав стверджувати про прямий причинно-наслідковий вплив лексики ризику на динаміку VIX.

Внесок авторів. *Автор повністю провів дослідження, написав і відредагував текст.*

Фінансування. *Автор заявляє, що дослідження виконано без зовнішнього фінансування.*

Конфлікт інтересів. *Автор заявляє про відсутність конфлікту інтересів.*

Додаток А. Робочий словник лексики ризику та невизначеності

Для виявлення лексики ризику та невизначеності використовувався розширений робочий словник, сформований за тематичним принципом із включенням синонімічних рядів і близьких за значенням лексичних одиниць. Такий підхід дає змогу враховувати різні мовні способи вираження ризику, нестабільності, фінансових втрат, загроз і невизначеності у фінансових новинах.

Базовий словник лексичних одиниць: risk; risky; risk-related; uncertainty; uncertain; unsure; volatile; volatility; fluctuate; fluctuation; instability; unstable; unpredictable; unpredictability; ambiguity; ambiguous; doubt; doubtful; speculation; speculative; threat; threaten; threatening; danger; dangerous; exposure; vulnerability; vulnerable; downside; loss; losses; decline; drop; fall; decrease; downturn; slump; crisis; crises; turmoil; turbulence; disruption; shock; pressure; stress; distress; default; failure; bankruptcy; insolvency; collapse; recession; slowdown; weakness; deterioration; adverse; negative; concern; concern-related; fear; fear-driven; warning; caution; cautionary.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Loughran T., McDonald B. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. *The Journal of Finance*. 2011. Vol. 66, No. 1. P. 35–65. DOI: https://doi.org/10.1111/j.1540-6261.2010.01625.x (дата звернення: 24.09.2026).
2. Bollen J., Mao H., Zeng X. Twitter mood predicts the stock market. *Journal of Computational Science*. 2011. Vol. 2, No. 1. P. 1–8. DOI: https://doi.org/10.1016/j.jocs.2010.12.007 (дата звернення: 24.09.2026).
3. Lachanski M., Pav S. Shy of the Character Limit: “Twitter Mood Predicts the Stock Market” Revisited. *Econ Journal Watch*. 2017. Vol. 14, No. 3. P. 302–345. URL: https://econjwatch.org/articles/shy-of-the-character-limit-twitter-mood-predicts-the-stock-market-revisited (дата звернення: 24.09.2026).
4. Smales L. A. Time-varying relationship of news sentiment, implied volatility and stock returns. *Applied Economics*. 2016. Vol. 48, No. 51. P. 4942–4960. DOI: https://doi.org/10.1080/00036846.2016.1167830 (дата звернення: 24.09.2026).
5. Kanelis D., Siklos P. L. The ECB Press Conference Statement: Deriving a New Sentiment Indicator for the Euro Area. *International Journal of Finance & Economics*. 2025. Vol. 30, No. 1. P. 652–664. DOI: https://doi.org/10.1002/ijfe.2940 (дата звернення: 24.09.2026).
6. Chau F., Deesomsak R., Shaikh R. Does Fed communication affect uncertainty and risk aversion? *Review of Quantitative Finance and Accounting*. 2025. Vol. 64. P. 713–756. DOI: https://doi.org/10.1007/s11156-024-01318-9 (дата звернення: 24.09.2026).
7. Poutachidou N., Bardas C., Papadamou S., Vlachos V. *Central Bank Communication Channel: International Evidence from Stock Market Indices*. 2025. 45 p. URL: https://ssrn.com/abstract=5252273 (дата звернення: 24.09.2026). DOI: https://doi.org/10.2139/ssrn.5252273 (дата звернення: 24.09.2026).

8. Kirtac K., Germano G. Sentiment trading with large language models. *Finance Research Letters*. 2024. Vol. 62. 105227. DOI: https://doi.org/10.1016/j.frl.2024.105227 (дата звернення: 24.09.2026).

9. Kadiyala L. A., Mirzaeinia A. Mamba Outpaces Reformer in Stock *Prediction with Sentiments from Top Ten LLMs*. 2025. DOI: https://doi.org/10.48550/arXiv.2510.01203 (дата звернення: 24.09.2026).

10. Yu B. Benchmarking Large Language Model Volatility. 2023. URL: https://arxiv.org/abs/2311.15180 (дата звернення: 24.09.2026).

11. Mino D., Williamson C. Sentiment and Volatility in Financial Markets: A Review of BERT and GARCH Applications during Geopolitical Crises. 2025. URL: https://arxiv.org/abs/2510.16503 (дата звернення: 24.09.2026).

12. Babii A., Ghysels E., Striaukas J. High-Dimensional Granger Causality Tests with an Application to VIX and News. *Journal of Financial Econometrics*. 2024. Vol. 22, No. 3. P. 605–635. DOI: https://doi.org/10.1093/jjfinec/nbac023 (дата звернення: 24.09.2026).

13. Tetlock P. C. Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*. 2007. Vol. 62, No. 3. P. 1139–1168. DOI: https://doi.org/10.1111/j.1540-6261.2007.01232.x (дата звернення: 24.09.2026).

14. Baker M., Wurgler J. Investor Sentiment in the Stock Market. *Journal of Economic Perspectives*. 2007. Vol. 21, No. 2. P. 129–151. DOI: https://doi.org/10.1257/jep.21.2.129 (дата звернення: 24.09.2026).

15. Garcia D. Sentiment during Recessions. *The Journal of Finance*. 2013. Vol. 68, No. 3. P. 1267–1300. DOI: https://doi.org/10.1111/j.1540-6261.2012.01805.x (дата звернення: 24.09.2026).

16. Yukhymenko T., Sorochan O. Impact of the Central Bank's Communications on FX Market Dynamics. *Visnyk of the National Bank of Ukraine*. 2023. No. 255. P. 4–21. URL: https://visnyk.bank.gov.ua/uploads/issues/255_en.pdf (дата звернення: 24.09.2026).

17. Szyszko M., Rutkowska A., Motuzka O. Central bank communication during the war: the case of the National Bank of Ukraine. *Post-Soviet Affairs*. 2025. Vol. 41, No. 6. P. 611–642. DOI: https://doi.org/10.1080/1060586X.2025.2528508 (дата звернення: 24.09.2026).

18. Benayoun D. Processed Dataset of Financial News Articles from Reuters (2006–2013). *Hugging Face Datasets*. URL: https://huggingface.co/datasets/danidanou/Reuters_Financial_News (дата звернення: 02.10.2026).

19. Financial Times. Markets News. URL: https://www.ft.com/markets (дата звернення: 02.10.2026).

20. Cboe Global Markets. Historical Data for Cboe VIX® Index and Other Volatility Indices. URL: https://www.cboe.com/tradable_products/vix/vix_historical_data (дата звернення: 02.10.2026).

REFERENCES

1. Loughran T., & McDonald B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. The Journal of Finance, vol. 66, no. 1, pp. 35–65. DOI: https://doi.org/10.1111/j.1540-6261.2010.01625.x (accessed September 24, 2026).
2. Bollen J., Mao H., & Zeng X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, vol. 2, no. 1, pp. 1–8. DOI: https://doi.org/10.1016/j.jocs.2010.12.007 (accessed September 24, 2026).
3. Lachanski M., & Pav S. (2017). Shy of the character limit: "Twitter mood predicts the stock market" revisited. Econ Journal Watch, vol. 14, no. 3, pp. 302–345. Available at: https://econjwatch.org/articles/shy-of-the-character-limit-twitter-mood-predicts-the-stock-market-revisited (accessed September 24, 2026).
4. Smales L. A. (2016). Time-varying relationship of news sentiment, implied volatility and stock returns. Applied Economics, vol. 48, no. 51, pp. 4942–4960. DOI: https://doi.org/10.1080/00036846.2016.1167830 (accessed September 24, 2026).
5. Kanelis D., & Siklos P. L. (2025). The ECB press conference statement: Deriving a new sentiment indicator for the euro area. International Journal of Finance & Economics, vol. 30, no. 1, pp. 652–664. DOI: https://doi.org/10.1002/ijfe.2940 (accessed September 24, 2026).
6. Chau F., Deesomsak R., & Shaikh R. (2025). Does Fed communication affect uncertainty and risk aversion? Review of Quantitative Finance and Accounting, vol. 64, pp. 713–756. DOI: https://doi.org/10.1007/s11156-024-01318-9 (accessed September 24, 2026).
7. Poutachidou N., Bardas C., Papadamou S., & Vlachos V. (2025). Central bank communication channel: International evidence from stock market indices, 45 p. DOI: https://doi.org/10.2139/ssrn.5252273 (accessed September 24, 2026).
8. Kirtac K., & Germano G. (2024). Sentiment trading with large language models. Finance Research Letters, vol. 62, art. 105227. DOI: https://doi.org/10.1016/j.frl.2024.105227 (accessed September 24, 2026).
9. Kadiyala L. A., & Mirzaeinia A. (2025). Mamba outpaces Reformer in stock prediction with sentiments from top ten LLMs. DOI: https://doi.org/10.48550/arXiv.2510.01203 (accessed September 24, 2026).
10. Yu B. (2023). Benchmarking large language model volatility. Available at: https://arxiv.org/abs/2311.15180 (accessed September 24, 2026).
11. Mino D., & Williamson C. (2025). Sentiment and volatility in financial markets: A review of BERT and GARCH applications during geopolitical crises. Available at: https://arxiv.org/abs/2510.16503 (accessed September 24, 2026).
12. Babii A., Ghysels E., & Striaukas J. (2024). High-dimensional Granger causality tests with an application to VIX and news. Journal of Financial Econometrics, vol. 22, no. 3, pp. 605–635. DOI: https://doi.org/10.1093/jjfinec/nbac023 (accessed September 24, 2026).
13. Tetlock P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. The Journal of Finance, vol. 62, no. 3, pp. 1139–1168. DOI: https://doi.org/10.1111/j.1540-6261.2007.01232.x (accessed September 24, 2026).

14. Baker M., & Wurgler J. (2007). Investor sentiment in the stock market. Journal of Economic Perspectives, vol. 21, no. 2, pp. 129–151. DOI: https://doi.org/10.1257/jep.21.2.129 (accessed September 24, 2026).

15. Garcia D. (2013). Sentiment during recessions. The Journal of Finance, vol. 68, no. 3, pp. 1267–1300. DOI: https://doi.org/10.1111/j.1540-6261.2012.01805.x (accessed September 24, 2026).

16. Yukhymenko T., & Sorochan O. (2023). Impact of the central bank's communications on FX market dynamics. Visnyk of the National Bank of Ukraine, no. 255, pp. 4–21. Available at: https://visnyk.bank.gov.ua/uploads/issues/255_en.pdf (accessed September 24, 2026).

17. Szyszko M., Rutkowska A., & Motuzka O. (2025). Central bank communication during the war: The case of the National Bank of Ukraine. Post-Soviet Affairs, vol. 41, no. 6, pp. 611–642. DOI: https://doi.org/10.1080/1060586X.2025.2528508 (accessed September 24, 2026).

18. Benayoun D. (n.d.). Processed dataset of financial news articles from Reuters (2006–2013). Hugging Face Datasets. Available at: https://huggingface.co/datasets/danidanou/Reuters_Financial_News (accessed October 2, 2026).

19. Financial Times. (n.d.). Markets news. Available at: https://www.ft.com/markets (accessed October 2, 2026).

20. Cboe Global Markets. (n.d.). Historical data for Cboe VIX® Index and other volatility indices. Available at: https://www.cboe.com/tradable_products/vix/vix_historical_data (accessed October 2, 2026).