

Л. О. ЛАТАНСЬКА, А. В. ПОЛОВИНКА  
Національний університет кораблебудування імені адмірала Макарова

## РЕГРЕСІЙНІ МОДЕЛІ ДЛЯ ОЦІНЮВАННЯ ТРИВАЛОСТІ РОЗРОБКИ БАНКІВСЬКОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ

*Розглянуто проблему підвищення достовірності оцінювання тривалості розробки ПЗ, що розробляється у щільно інтегрованому середовищі банківських інформаційних систем, яка є критично важливою для ефективного планування проєктів, оптимального розподілу ресурсів та забезпечення їх комерційної успішності. Показано, що існуючі підходи до оцінювання термінів розробки ПЗ, зокрема моделі COCOMO і ISBSG, не забезпечують достатньої достовірності для специфічних умов банківської сфери, що характеризується високою складністю архітектури, інтеграцією численних зовнішніх сервісів, регуляторними вимогами та підвищеними вимогами до безпеки.*

*Спираючись на дані попередніх досліджень та результати аналізу, обґрунтовано доцільність використання нелінійних регресійних моделей із попередньою нормалізацією даних для оцінювання тривалості розробки банківського ПЗ на основі трудомісткості проєкту. Запропоновано підхід, який передбачає побудову окремих моделей для однокомандних і багатоконандних проєктів, що дозволяє врахувати додаткові витрати часу, пов'язані з міжкомандною взаємодією, синхронізацією робіт і організаційною складністю. Емпіричну основу дослідження склали дані 482 реальних проєктів розробки банківського ПЗ за 2014–2024 роки, створених з використанням мов програмування Java і JavaScript.*

*Реалізовано методику побудови нелінійних регресійних моделей, яка включає нормалізацію даних із використанням перетворень десятичного логарифму, двовимірного перетворення Бокса-Кокса та одновимірного перетворення Джонсона, виявлення та вилучення викидів на основі еліпсу прогнозування, побудову лінійної регресії для нормалізованих даних, отримання нелінійних моделей шляхом зворотних перетворень, побудову довірчих інтервалів та інтервалів прогнозування, а також пошук викидів за допомогою інтервалів прогнозування. Якість моделей перевірено за допомогою показників якості  $R^2$ , MMRE та PRED(0,25).*

*Отримані результати свідчать, що моделі, побудовані з використанням логарифмічного перетворення, є непридатними через значну кількість викидів. Натомість моделі на основі перетворень Бокса-Кокса та Джонсона демонструють прийнятну достовірність: значення  $R^2$  перевищують 0,85, MMRE не перевищує 0,25, а PRED(0,25) для більшості випадків вище 0,75. Встановлено, що найкращі результати забезпечують моделі з використанням двовимірного перетворення Бокса-Кокса, які характеризуються меншою кількістю викидів та кращими показниками узгодженості з емпіричними даними.*

*Запропоновані моделі можуть бути використані для підвищення достовірності прогнозування тривалості розробки банківського ПЗ, створеного за допомогою мов програмування Java і JavaScript, на основі трудомісткості проєктів. Водночас визначено обмеження підходу, зокрема залежність моделей від специфіки предметної області та обмеження діапазону значень трудомісткості проєктів.*

**Ключові слова:** нелінійна регресійна модель, банківське ПЗ, тривалість розробки ПЗ, трудомісткість проєкту, оцінювання, викид, нормалізуюче перетворення, перетворення Бокса-Кокса, перетворення Джонсона.

L. O. LATANSKA, A. V. POLOVYNKA  
Admiral Makarov National University of Shipbuilding

## REGRESSION MODELS FOR ESTIMATING THE DURATION OF BANKING SOFTWARE DEVELOPMENT

*The paper addresses the problem of improving the estimation accuracy of software development duration in a tightly integrated environment of banking information systems, which is critically important for effective project planning, optimal resource allocation and ensuring business success. It is shown that existing approaches to estimating software development time, in particular the COCOMO and ISBSG models, do not provide sufficient accuracy for the specific conditions of the banking domain, which is characterized by high architectural complexity, integration of numerous external services, regulatory requirements and increased security demands.*

*Based on data from previous studies and analysis results, the use of nonlinear regression models with prior data normalization for estimating the duration of banking software development based on project effort is justified. An approach is proposed that involves constructing separate models for single-team and multi-team projects, which allows the model to account for additional time costs due to inter-team interaction, work synchronization and organizational complexity. The empirical basis of the study consists of data from 482 real banking software development projects for the period 2014–2024, developed in Java and JavaScript.*

*A methodology for constructing nonlinear regression models is implemented, which includes data normalization using decimal logarithm transformations, the bivariate Box-Cox transformation and the univariate Johnson transformation, detection and removal of outliers based on the prediction ellipse, construction of linear regression for normalized data, derivation of nonlinear models through inverse transformations, construction of confidence intervals and prediction intervals and identification of outliers using prediction intervals. The quality of the models is evaluated using the  $R^2$ , MMRE and PRED(0.25) metrics.*

*The obtained results indicate that models constructed using logarithmic transformation are unsuitable due to a significant number of outliers. In contrast, models based on the Box-Cox and Johnson transformations demonstrate acceptable accuracy:  $R^2$  values exceed 0.85, MMRE does not exceed 0.25 and PRED(0.25) is above 0.75 in most cases.*

*The best results are achieved by models using the bivariate Box-Cox transformation, which are characterized by a smaller number of outliers and better agreement with empirical data.*

*The proposed models can be used to improve the accuracy of predicting the duration of banking software development in Java and JavaScript based on project effort. At the same time, the limitations of the approach are identified, in particular the dependence of the models on the specifics of the application domain and the limitation of the range of project effort values.*

**Keywords:** *nonlinear regression model, banking software, software development duration, project effort, estimation, outlier, normalizing transformation, Box-Cox transformation, Johnson transformation.*

### Постановка проблеми

Відповідно до досліджень у сфері управління програмними проектами, порушення термінів реалізації є найпоширенішим видом невдач, з якими стикаються проєктні команди [1]. У більшості випадків на це впливають не тільки технічні чинники – причини також можуть бути організаційними або комунікаційними. Оскільки успішність розробки проєкту напряму впливає на комерційні та репутаційні результати компанії або фінансової установи, то достовірність оцінювання тривалості створення проєкту є одним із ключових чинників в сфері розробки ПЗ, який безпосередньо використовується в плануванні проєкту та визначенні доцільності його реалізації.

Проєкти з розробки банківського ПЗ мають особливості, що суттєво ускладнюють точне оцінювання термінів їх реалізації. З поміж основних виділяють складність регуляторних вимог, складність архітектури ПЗ та інфраструктури, виняткову щільність пов'язаних між собою інформаційних потоків, що призводить до необхідності внесення змін в багатьох інформаційних системах в рамках реалізації одного проєкту. Також на фінальні строки проєкту впливають високі вимоги до безпеки та більш задокументований процес прийняття рішень у порівнянні з комерційним ПЗ.

Труднощі оцінювання тривалості виконання робіт з розробки ПЗ зумовлені постійним зростанням складності програмних компонентів та залежностей між різними модулями ПЗ, особливо при використанні мікросервісної архітектури розробки, які необхідно відповідно врахувати. Тож задача достовірного оцінювання термінів розробки банківського ПЗ є актуальною, оскільки її вирішення напряму впливає на правильне визначення пріоритетів планування, ефективне використання ресурсів та, в результаті, на комерційний успіх проєкту.

### Аналіз останніх досліджень та публікацій

Основні ризики, що впливають на строки виконання проєкту, можна згрупувати за кількома категоріями:

– недостатня достовірність оцінювання строків. Експертне оцінювання залишається домінуючим методом, проте його суб'єктивність часто веде до недооцінки складності. Систематичні помилки в оцінюванні часу є одним із ключових чинників, що призводять до недотримання графіків та остаточного провалу програмних проєктів [2];

– неповнота або зміна вимог. Це змушує команди витратити значну частину часу на переробку вже готового функціоналу. Нестабільність вимог має суттєвий вплив на перевищення строків виконання проєктів та призводить до затримок і перевитрат ресурсів [3–5];

– недостатній рівень планування та управління ризиками. Відсутність резервів часу, формалізованого управління ризиками та деталізованого планування призводить до підвищеної чутливості проєкту до невизначеностей і, як наслідок, до порушення строків його виконання [4];

– технічна або організаційна складність [5; 6]. Сучасні системи часто мають розподілену архітектуру, численні інтеграції та залежності. Найменші затримки у суміжних модулях можуть кардинально впливати на критичний шлях проєкту і, відповідно, на його тривалість;

– обмежені компетенції та ресурси команди. Недостатність кваліфікованих кадрів або багатозадачність серед учасників команди негативно впливають на продуктивність і терміни завершення продукту [7];

– комунікаційні та координаційні бар'єри в команді, зокрема у розподілених проєктах, можуть негативно впливати на ефективність виконання робіт і призводити до затримок [4; 6].

Проведений аналіз сучасних підходів до оцінювання тривалості розробки ПЗ показав що, незважаючи на широке застосування методів машинного навчання [8], нелінійні регресійні моделі СОСОМО або їх модифікації [9; 10] та моделі на основі даних ISBSG, у яких тривалість проєкту визначається як функція його трудомісткості, продовжують активно використовуватись. Однак застосування даних моделей для існуючого набору емпіричних даних реальних проєктів з розробки банківського ПЗ показало, що якість отриманих моделей СОСОМО та ISBSG недостатня для достовірного прогнозування [11]. Також у роботі [11] було підтверджено, що обраний набір емпіричних даних не відповідає багатовимірному нормальному закону розподілу, що не дозволяє побудувати достовірну лінійну регресійну модель.

У результаті була визначена потреба у побудові нелінійних регресійних моделей з різними нормалізуючими перетвореннями для оцінювання тривалості проєктів з розробки банківського ПЗ.

Враховуючи те, що основним фактором у більшості існуючих моделей для оцінювання тривалості проєкту розробки ПЗ є трудомісткість, побудову відповідних моделей ми також здійснюємо в залежності від зазначеної метрики.

Побудова визначеної моделі за наявними емпіричними даними без диференціації проєктів за окремими критеріями призвела до негативного результату, оскільки значна частина даних була визначена як викиди, а достовірність моделі виявилась незадовільною [12].

У результаті подальшого аналізу даних було сформовано гіпотезу, що причиною неможливості побудувати єдину достовірну модель для емпіричних даних виступає значна різниця між тривалістю однокомандних і багатокомандних проєктів. Така різниця зумовлена появою суттєвих додаткових комунікаційних та методологічних витрат, пов'язаних із взаємодією між командами, несинхронністю графіків видачі версій різними командами, різницею у їх залученості та пріоритетах задач, появою додаткових ролей (проєктного менеджера або скрам-майстра), а також потребою побудови окремої платформи для спільного інтеграційного тестування.

Отже, щоб отримати достовірні моделі для оцінювання тривалості розробки банківського ПЗ, ми створюємо окремі моделі для однокомандних і багатокомандних проєктів.

### Мета дослідження

Метою дослідження є побудова двох наборів нелінійних регресійних моделей для оцінювання тривалості розробки банківського ПЗ, що створене з застосуванням мов програмування Java та JavaScript, для однокомандних і багатокомандних проєктів з використанням різних нормалізуючих перетворень.

Об'єктом дослідження є процес оцінювання тривалості розробки ПЗ, що розробляється у щільно інтегрованому середовищі банківських інформаційних систем, побудованих із використанням мов програмування Java та JavaScript.

Предметом дослідження є нелінійні регресійні моделі для оцінювання тривалості розробки ПЗ, що розробляється у щільно інтегрованому середовищі банківських інформаційних систем, яке створюється із застосуванням мов програмування Java та JavaScript.

Для досягнення поставленої мети потрібно вирішити наступні завдання:

- Побудувати два набори нелінійних регресійних моделей для оцінювання тривалості однокомандних і багатокомандних проєктів розробки банківського ПЗ із застосуванням різних нормалізуючих перетворень, а саме: десяткового логарифму, двовимірного перетворення Бокса-Кокса та одновимірного перетворення Джонсона.
- Перевірити якість побудованих регресійних моделей.
- Виконати аналіз результатів та визначити потенційні шляхи проведення подальших досліджень з оцінювання тривалості проєктів з розробки банківського ПЗ.

### Виклад основного матеріалу дослідження

В якості емпіричних даних для побудови регресійних моделей оцінювання тривалості проєктів з розробки банківського ПЗ були взяті дані з роботи [11]. Вони включали інформацію по 482 реальним проєктам розробки банківського ПЗ за 2014–2024 рр, створеним за допомогою мов програмування Java (backend-розробка) і JavaScript (frontend-розробка).

Для побудови моделей були використані дані по тривалості та трудомісткості виконаних проєктів, а також дані по кількості команд, задіяних у проєкті. В залежності від кількості команд розробників проєкти були розподілені на дві групи: однокомандні проєкти та проєкти із залученням більшої кількості команд. Кожна з цих груп додатково була розділена на навчальну та тестову вибірки у співвідношенні 60 % та 40 % первинної вибірки. При цьому було забезпечено наявність у навчальній вибірці мінімального та максимального значення метрик, а у тестовій вибірці – значень, близьких або рівних граничним. Побудова моделей відбувалась за даними навчальних вибірок. Кількісні результати розподілу наведено в таблиці 1.

Таблиця 1

#### Отримана кількість проєктів за результатами розподілу емпіричних даних за критерієм кількості команд

Група проєктів	Загальна кількість	Навчальна вибірка	Тестова вибірка
Однокомандні проєкти	266	177	89
Багатокомандні проєкти	216	144	72

Перевірка емпіричних даних з тривалості та трудомісткості проєктів на багатовимірну нормальність розподілу була виконана за допомогою критеріїв Мардіа. В результаті розподіли цих двовимірних даних для обох груп проєктів не відповідають нормальному розподілу: значення двовимірних асиметрії та ексцесу перевищують значення відповідних тестових статистик [11].

Для побудови нелінійних регресійних моделей було використано методику, наведену в [13; 14]. На першому етапі була виконана нормалізація емпіричних даних. Для нормалізації даних були використані: нормалізуюче перетворення за десятковим логарифмом, двовимірне перетворення Бокса-Кокса, одновимірне перетворення Джонсона. Для кожного перетворення були побудовані окремі моделі, що дозволило провести порівняльний аналіз і обрати оптимальну модель.

Для оцінювання ефективності перетворення після нормалізації отриманий набір нормалізованих даних також було перевірено на двовимірну нормальність за допомогою критеріїв Мардіа, щоб підтвердити готовність даних для подальшого аналізу з використанням статистичних методів, розроблених для нормального розподілу.

Наступним етапом обробки даних було виявлення та очищення даних від викидів. Їх ідентифікація та видалення є необхідними, оскільки вони можуть суттєво спотворювати результати аналізу та знижувати достовірність моделі. Пошук викидів в отриманих нормалізованих двовимірних даних було здійснено за допомогою еліпсу прогнозування, рівняння якого може бути записано як:

$$(Z - \bar{Z})^T S_N^{-1} (Z - \bar{Z}) = \frac{2(N^2 - 1)}{N(N - 2)} F_{2, N-2, \alpha}, \quad (1)$$

де ліва частина рівняння є квадратом відстані Махаланобіса;  $F_{2, N-2, \alpha}$  – квантиль  $F$ -розподілу з 2 та  $N - 2$  ступенями свободи;  $N$  – кількість точок даних;  $\alpha$  – рівень значущості (було використано  $\alpha = 0,005$ );  $\bar{Z}$  – вектор вибірових середніх;  $S_N^{-1}$  – обернена вибіркова коваріаційна матриця.

Вихід точки нормалізованих даних за границі еліпсу прогнозування означає, що ця точка є двовимірним викидом. У разі її знаходження і видалення з набору даних виконувалось повернення на перший крок і нормалізація оновлених даних.

Після очищення даних були побудовані лінійні регресійні моделі для нормалізованих двовимірних даних у вигляді

$$Z_Y = \hat{Z}_Y + \varepsilon = \hat{b}_0 + \hat{b}_1 \cdot Z_1 + \varepsilon, \quad (2)$$

де  $\hat{Z}_Y$  є результатом передбачення за рівнянням лінійної регресії;  $\varepsilon$  – гаусівська випадкова величина, що визначає залишки,  $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ ;  $\hat{b}_0, \hat{b}_1$  – оцінки параметрів регресії.

Однією із умов застосування лінійної регресії є нормальність розподілу відхилення  $\varepsilon$ , тож далі за допомогою критерію  $\chi^2$  Пірсона перевірялась нормальність розподілу залишків  $\varepsilon$  для отриманої за (2) лінійної регресійної моделі для нормалізованих даних. Якщо розподіл не був гаусівським, вилучався рядок двовимірних даних з максимальним модулем залишку і процес повторювався з першого етапу.

Наступним кроком була побудова нелінійної регресійної моделі:

$$\hat{Y} = \psi_Y^{-1}(\hat{Z}_Y + \varepsilon) = \psi_Y^{-1}(\hat{b}_0 + \hat{b}_1 \cdot Z_1 + \varepsilon), \quad (3)$$

де  $\hat{Y}$  – оцінка тривалості розробки проекту,  $\psi_Y$  – перша компонента вектору  $\psi$  перетворення,  $Z_1$  – вектор нормалізованих значень трудомісткості проектів,  $\hat{b}_0, \hat{b}_1$  – оцінки параметрів лінійної регресійної моделі нормалізованих даних, які були отримані за допомогою методу найменших квадратів.

У якості  $\psi_Y^{-1}$  були використані зворотні перетворення, відповідно до використаних нормалізуючих: перетворення за десятковим логарифмом, Бокса-Кокса та Джонсона.

На останньому кроці було побудовано довірчі інтервали та інтервали прогнозування нелінійної регресії та виконано перевірку на наявність точок початкового розподілу, які виходять за межі інтервалу прогнозування. Якщо такі точки були знайдені, то вони вилучалися і процес повторювався з першого етапу. Довірчий інтервал нелінійної регресії визначався за формулою:

$$\psi_Y^{-1} \left( \hat{Z}_Y \pm t_{\alpha/2, \nu} S_{Z_Y} \left\{ \frac{1}{N} + (z_X^+)^T S_{ZZ}^{-1} (z_X^+) \right\}^{1/2} \right), \quad (4)$$

а інтервал прогнозування за формулою [13]:

$$\psi_Y^{-1} \left( \hat{Z}_Y \pm t_{\alpha/2, \nu} S_{Z_Y} \left\{ 1 + \frac{1}{N} + (z_X^+)^T S_{ZZ}^{-1} (z_X^+) \right\}^{1/2} \right), \quad (5)$$

де  $\hat{Z}_y$  – результат прогнозування за лінійним регресійним рівнянням (2) для нормалізованих даних;  $t_{\alpha/2, v}$  – квантиль розподілу Стьюдента з рівнем значущості  $\alpha/2$  та  $v$  ступенями свободи;  $v = N - k - 1$ , де  $k$  – кількість незалежних змінних (в нашому випадку  $k = 1$ );  $z_X^+$  – вектор центрованих нормалізованих факторів, який містить значення  $Z_{1_i} - \bar{Z}_1, Z_{2_i} - \bar{Z}_2, \dots, Z_{k_i} - \bar{Z}_k$ ;  $S_{ZZ}$  – матриця розмірності  $k \times k$ :

$$S_{ZZ} = \begin{bmatrix} S_{Z_1Z_1} & S_{Z_1Z_2} & \dots & S_{Z_1Z_k} \\ S_{Z_2Z_1} & S_{Z_2Z_2} & \dots & S_{Z_2Z_k} \\ \dots & \dots & \dots & \dots \\ S_{Z_kZ_1} & S_{Z_kZ_2} & \dots & S_{Z_kZ_k} \end{bmatrix},$$

де  $S_{Z_qZ_r} = \sum_{i=1}^N [Z_{q_i} - \bar{Z}_q][Z_{r_i} - \bar{Z}_r]$ ,  $q, r = 1, 2, \dots, k$ .

Після пошуку викидів в отриманих нормалізованих двовимірних даних за допомогою (1), були побудовані нелінійні регресійні моделі за (3).

При побудові моделей за наведеним алгоритмом, з використанням логарифмічного нормалізуючого перетворення для однокомандних та багатокомандних проєктів, на основі навчальних вибірок з 177 та 144 точками даних відповідно, було отримано значну кількість викидів – 110 для однокомандних проєктів та 71 для багатокомандних, що складає 62 % та 49 % відповідно). З цього було зроблено висновок про те, що модель, побудована на залишку даних, не зможе достовірно описати процес, що досліджується, і про недоцільність подальшої побудови нелінійної регресійної моделі на основі даного перетворення.

Кількість викидів, значення параметрів перетворення та оцінки параметрів моделей, побудованих для однокомандних та багатокомандних проєктів з використанням нормалізуючих перетворень Бокса-Кокса та Джонсона наведені відповідно у таблицях 2 і 3.

Таблиця 2

**Результати побудови моделей з використанням нормалізуючого перетворення Бокса-Кокса**

№ моделі	Група проєктів	Кількість викидів	Параметри перетворення для X	Параметри перетворення для Y	$\hat{b}_0$	$\hat{b}_1$
1	Одно-командні	37	$1 = -0,2851$	$1 = -0,5700$	1,0515	0,2233
2	Багато-командні	27	$1 = 0,2017$	$1 = -0,0101$	2,2450	0,2835

Таблиця 3

**Результати побудови моделей з використанням нормалізуючого перетворення Джонсона**

№ моделі	Група проєктів	Кількість викидів	Параметри перетворення для X	Параметри перетворення для Y	$\hat{b}_0$	$\hat{b}_1$
3	Одно-командні	51	$\gamma = 1,333$ $v = 0,7142$ $\phi = 46,7089$ $\lambda = 374,4019$	$\gamma = 1,8873$ $v = 0,9775$ $\phi = 46,1336$ $\lambda = 426,1221$	$3 \cdot 10^{-6}$	0,9474
4	Багато-командні	40	$\gamma = 18,1915$ $v = 1,3724$ $\phi = 47,9129$ $\lambda = 1,4432 \cdot 10^8$	$\gamma = 14,9352$ $v = 1,1771$ $\phi = 72,1426$ $\lambda = 5,2873 \cdot 10^7$	$5 \cdot 10^{-6}$	0,9814

Обчислені за формулами (4) та (5) ширини довірчих інтервалів та інтервалів прогнозування нелінійної регресії на основі двовимірного нормалізуючого перетворення Бокса-Кокса для більшості точок даних виявилися найменшими.

Лінії нелінійних регресій, довірчі інтервали та інтервали прогнозування нелінійних регресій для довірчої ймовірності 0,95 для моделей, побудованих на основі двовимірного нормалізуючого перетворення Бокса-Кокса, представлені на рисунках 1 (однокомандні проекти) та 2 (багатокомандні проекти). Зелені лінії на рисунках відповідають лініям регресій, сині лінії – довірчим інтервалам, помаранчеві – інтервалам прогнозування відповідних регресій.

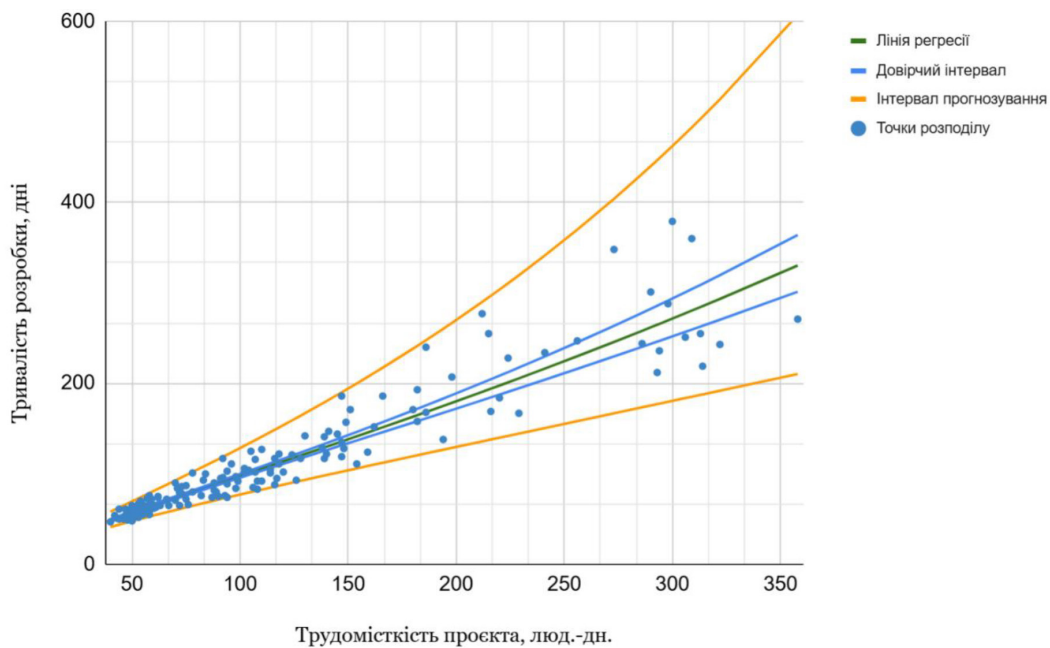


Рис. 1. Довірчі інтервали та інтервали прогнозування нелінійної регресії на основі нормалізуючого перетворення Бокса-Кокса для однокомандних проєктів

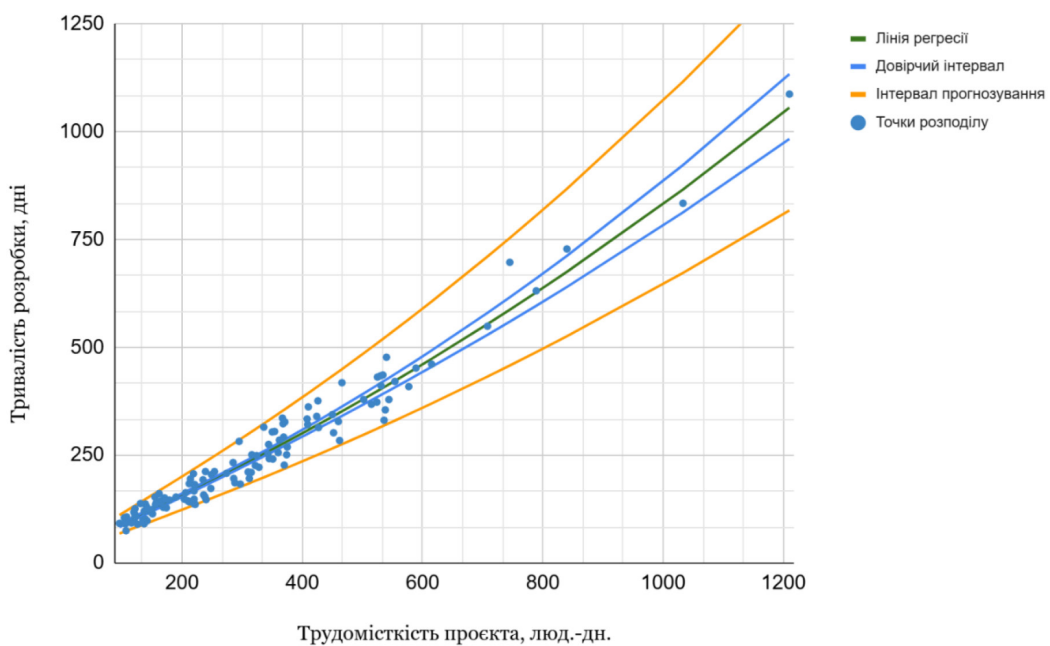


Рис. 2. Довірчі інтервали та інтервали прогнозування нелінійної регресії на основі нормалізуючого перетворення Бокса-Кокса для багатокомандних проєктів

Для оцінювання достовірності побудованих нелінійних регресійних моделей була перевірена їх якість шляхом розрахунку наступних параметрів: вибіркового коефіцієнту детермінації  $R^2$ , середньої величини відносної похибки  $MMRE$  та відсотка прогнозованих результатів, для яких величини відносної похибки менші за  $0,25 - PRED(0,25)$ .

Результати оцінювання якості моделей для навчальної та тестової вибірок наведені у таблиці 4.

Таблиця 4

**Показники якості побудованих моделей для навчальної та тестової вибірок**

Модель	Група проєктів	Вибірка	Кількість викидів	$R^2$	$MMRE$	$PRED(0,25)$
Модель 1 (Бокс-Кокс)	Одно-командні	Навчальна	37	0,8762	0,1129	0,9143 (128/140)
		Тестова	–	0,8697	0,1770	0,7416 (66/89)
Модель 2 (Бокс-Кокс)	Багато-командні	Навчальна	27	0,9654	0,1016	0,9658 (113/117)
		Тестова	–	0,9061	0,1608	0,7639 (55/72)
Модель 3 (Джонсон)	Одно-командні	Навчальна	51	0,8962	0,1108	0,9206 (116/126)
		Тестова	7*	0,8597	0,1710	0,7317 (60/82)
Модель 4 (Джонсон)	Багато-командні	Навчальна	40	0,978	0,0869	0,9808 (102/104)
		Тестова	–	0,8944	0,1744	0,6944 (55/72)

\* – частина точок з тестової вибірки не вклались в граничні параметри побудованої моделі (після видалення викидів мінімальне значення параметру  $X$  для моделі змінилось з 40 на 47, що призвело до вибраковки частини тестових даних)

Загалом результати побудови моделей з таблиці 4 демонструють їх задовільну якість: значення  $R^2$  більші ніж 0,85 вказують на те, що більшість значень  $Y$  розташовані близько до лінії регресії. Зазвичай прийнятною точністю прогнозування за допомогою регресійних моделей вважаються значення  $MMRE \leq 0,25$  та  $PRED(0,25) \geq 0,75$ . Згідно таблиці 4 всі побудовані моделі мають  $MMRE < 0,25$  та для навчальних вибірок  $PRED(0,25) > 0,9$ . Для тестових вибірок результат трохи гірший: для однієї моделі на базі перетворення Бокса-Кокса і двох моделей на базі перетворення Джонсона  $PRED(0,25) < 0,75$  на 1–6 %. Це свідчить про прийнятну якість побудованих нелінійних регресійних моделей і можливість їх використання для оцінювання тривалості розробки банківського ПЗ на основі його трудомісткості.

Також значення показників якості свідчать про кращу якість моделей, побудованих на основі двовимірного нормалізуючого перетворення Бокса-Кокса у порівнянні з одновимірним перетворенням Джонсона – параметри якості, як для навчальних, так і, зокрема, для тестових вибірок, кращі, особливо для однокомандних проєктів, де 7 точок тестової вибірки не вклалися у граничні параметри побудованої моделі. На наш погляд, це пояснюється меншою кількістю викидів, які були отримані при побудові моделей з перетворенням Бокса-Кокса, завдяки чому отримані моделі більше відповідають початковим даним і їх достовірність вище.

До обмежень побудованих нелінійних регресійних моделей можна віднести те, що вони побудовані для проєктів розробки банківського ПЗ і враховують саме їх особливості. Також моделі потребують знань щодо кількості команд розробки, які були задіяні у проєкті. Для моделі 1 трудомісткість розробки проєкта знаходиться у межах від 40 до 358 людиноднів, для моделі 2 – від 96 до 2080, для моделі 3 – від 48 до 358, для моделі 4 – у межах від 96 до 1210.

**Висновки**

У роботі були побудовані нелінійні регресійні моделі для оцінювання тривалості розробки банківського ПЗ в залежності від трудомісткості розробки з розділенням за критерієм кількості команд, що приймають в ній участь. Моделі на основі нормалізуючого перетворення десятичного логарифму були визнані непридатними по причині великої кількості викидів. Моделі,

що були побудовані на основі двовимірного нормалізуючого перетворення Бокса-Кокса та одновимірного перетворення Джонсона, в цілому демонструють задовільну якість оцінювання за показниками  $R^2$ ,  $MMRE$  та  $PRED(0,25)$ . При цьому моделі, побудовані на основі нормалізуючого перетворення Бокса-Кокса, показали меншу кількість отриманих викидів та кращі показники якості і рекомендуються для подальшого використання для оцінювання тривалості розробки банківського ПЗ на основі його трудомісткості.

Проте дані моделі теж мають певні недоліки: кількість викидів при побудові моделей була суттєвою, для однокомандних проєктів вона склала 20,9 % даних, а для багатокомандних – 18,8 %. Це призвело до зниження показників  $PRED(0,25)$  при аналізі якості тестової вибірки до 74–76 % у порівнянні з 91–96 % для відповідних навчальних вибірок.

Наукова новизна дослідження полягає в наступному. Вперше побудовано однофакторні нелінійні регресійні моделі з використанням двовимірного нормалізуючого перетворення Бокса-Кокса та одновимірного нормалізуючого перетворення Джонсона для оцінювання тривалості розробки ПЗ, що розробляється у щільно інтегрованому середовищі банківських інформаційних систем, побудованих з використанням мов програмування Java та JavaScript. Це, у порівнянні з існуючими моделями, дозволило підвищити достовірність оцінювання тривалості розробки: були отримані більші значення вибіркового коефіцієнту детермінації  $R^2$  і відсотка прогнозованих результатів  $PRED(0,25)$  та менші значення середньої величини відносної похибки  $MMRE$ .

Отримали подальший розвиток рівняння для визначення границь інтервалів прогнозування нелінійних регресійних моделей для оцінювання тривалості розробки банківського ПЗ, побудованого з використанням мов програмування Java та JavaScript, за рахунок використання двовимірного перетворення Бокса-Кокса, що дозволило зменшити ширину довірчих інтервалів та інтервалів прогнозування.

Практична цінність даної роботи полягає в можливості використання побудованих моделей для прогнозування тривалості проєктів, що розробляються у щільно інтегрованому середовищі банківських інформаційних систем, побудованих з використанням мов програмування Java та JavaScript, для покращення точності їх планування і оцінювання.

У подальшому планується для підвищення достовірності моделювання розглянути побудову багатфакторних нелінійних регресійних моделей, зокрема, з використанням додаткового фактору кількості розробників, задіяних на проєкті.

### Список використаної літератури

1. Mersino A. Project Managers Fail to Help Software Projects (Standish Group Chaos 2020). URL: <https://medium.com/leadership-and-agility/project-managers-fail-to-help-software-projects-standish-group-chaos-2020-e65d803e99f2> (дата звернення: 12.03.2025).
2. Pasuksmit J., Thongtanunam P., Karunasekera S. A Systematic Literature Review on Reasons and Approaches for Accurate Effort Estimations in Agile. *ACM Computing Surveys*. 2024. Vol. 56 (11). P. 1–37. URL: <https://doi.org/10.1145/3663365>
3. Dasanayake S., Aaramaa S., Markkula J., Oivo M. Impact of requirements volatility on software architecture: How do software teams keep up with ever-changing requirements? *Journal of Software: Evolution and Process*. 2019. Vol. 31(6). P. e2160 URL: <https://doi.org/10.1002/smr.2160>
4. Ashleigh M., Ojiako U., Chipulu M., Wang J. K. Critical learning themes in project management education: Implications for blended learning. *International Journal of Project Management*. 2012. Vol. 30 (2). P. 153–161. URL: <https://doi.org/10.1016/j.ijproman.2011.05.002>
5. Morgenshtern O., Raz T., Dvir D. Factors affecting duration and effort estimation errors in software development projects. *Information and Software Technology*. 2007. Vol. 49(8). P. 827–837, URL: <https://doi.org/10.1016/j.infsof.2006.09.006>

6. Teslyuk V., Batyuk A., Voityshyn V. Method of Software Development Project Duration Estimation for Scrum Teams with Differentiated Specializations. *Systems*. 2022. Vol 10(4). Article 123. URL: <https://doi.org/10.3390/systems10040123>
7. Korytkowski P., Malachowski B. Competence-based estimation of activity duration in IT projects. *European Journal of Operational Research*. 2019. Vol. 275 (2). P. 708–720. URL: <https://doi.org/10.1016/j.ejor.2018.11.061>
8. Pospieszny P., Czarnacka-Chrobot B., Kobylinski A. An effective approach for software project effort and duration estimation with machine learning algorithms. *Journal of Systems and Software*. 2018. Vol. 137. P. 184–196. URL: <https://doi.org/10.1016/j.jss.2017.11.066>
9. Menzies T., Yang Y., Mathew G., Boehm B., Hihn J. Negative results for software effort estimation. *Empir Software Eng*. 2017. Vol. 22, P. 2658–2683. URL: <https://doi.org/10.1007/s10664-016-9472-2>
10. Nandal D., Sangwan O. P. Software cost estimation by optimizing COCOMO model using hybrid BATGSA algorithm. *International Journal of Intelligent Engineering and Systems*. 2018. 11(4). P. 250–263. URL: <https://doi.org/10.22266/ijies2018.0831.25>
11. Latanska L. O., Polovynka A. V. Analysis of methods and models for estimating the duration of banking software development projects. *Herald of Khmelnytskyi national university. Technical sciences*. 2025. Issue 3, part 2. P.140–145. <https://doi.org/10.31891/2307-5732-2025-353-18>
12. Латанська Л. О., Половинка А. В. Математичне моделювання в задачах оцінювання тривалості розробки банківського програмного забезпечення. *Комп'ютерне моделювання та оптимізація складних систем (КМОСС-2025) : матеріали ІХ Міжнародної науково-технічної конференції, м. Дніпро, 5–7 листопада 2025 р. Дніпро, 2025. С. 59.*
13. Prykhodko N. V., Prykhodko S. B. Constructing the Nonlinear Regression Models on the Basis of Multivariate Normalizing Transformations. *Electronic Modeling*. 2018. Vol. 40. No 6. P. 99–108. URL: <https://doi.org/10.15407/emodel.40.06.101>
14. Prykhodko S. B., Prykhodko N. V., Koltsov A. V. A Nonlinear Regression Model for Early LOC Estimation of Open-Source Kotlin-Based Applications. *Radio Electronics Computer Science Control*. 2024. No 1. P. 85–95. URL: <https://doi.org/10.15588/1607-3274-2024-1-8>

### References

1. Mersino, A. (2020). *Project managers fail to help software projects (Standish Group Chaos 2020)*. Medium. <https://medium.com/leadership-and-agility/project-managers-fail-to-help-software-projects-standish-group-chaos-2020-e65d803e99f2>
2. Pasuksmit, J., Thongtanunam, P., & Karunasekera, S. (2024). A systematic literature review on reasons and approaches for accurate effort estimations in agile. *ACM Computing Surveys*, 56(11), 1–37. <https://doi.org/10.1145/3663365>
3. Dasanayake, S., Aaramaa, S., Markkula, J., & Oivo, M. (2019). Impact of requirements volatility on software architecture: How do software teams keep up with ever-changing requirements? *Journal of Software: Evolution and Process*, 31(6), e2160. <https://doi.org/10.1002/smr.2160>
4. Ashleigh, M., Ojiako, U., Chipulu, M., & Wang, J. K. (2012). Critical learning themes in project management education: Implications for blended learning. *International Journal of Project Management*, 30(2), 153–161. <https://doi.org/10.1016/j.ijproman.2011.05.002>
5. Morgenstern, O., Raz, T., & Dvir, D. (2007). Factors affecting duration and effort estimation errors in software development projects. *Information and Software Technology*, 49(8), 827–837. <https://doi.org/10.1016/j.infsof.2006.09.006>
6. Teslyuk, V., Batyuk, A., & Voityshyn, V. (2022). Method of software development project duration estimation for Scrum teams with differentiated specializations. *Systems*, 10(4), 123. <https://doi.org/10.3390/systems10040123>

7. Korytkowski, P., & Małachowski, B. (2019). Competence-based estimation of activity duration in IT projects. *European Journal of Operational Research*, 275(2), 708–720. <https://doi.org/10.1016/j.ejor.2018.11.061>
8. Pospieszny, P., Czarnacka-Chrobot, B., & Kobylinski, A. (2018). An effective approach for software project effort and duration estimation with machine learning algorithms. *Journal of Systems and Software*, 137, 184–196. <https://doi.org/10.1016/j.jss.2017.11.066>
9. Menzies, T., Yang, Y., Mathew, G., Boehm, B., & Hihn, J. (2017). Negative results for software effort estimation. *Empirical Software Engineering*, 22, 2658–2683. <https://doi.org/10.1007/s10664-016-9472-2>
10. Nandal, D., & Sangwan, O. P. (2018). Software cost estimation by optimizing COCOMO model using hybrid BATGSA algorithm. *International Journal of Intelligent Engineering and Systems*, 11(4), 250–263. <https://doi.org/10.22266/ijies2018.0831.25>
11. Latanska, L. O., & Polovynka, A. V. (2025). Analysis of methods and models for estimating the duration of banking software development projects. *Herald of Khmelnytskyi National University. Technical Sciences*, 3(2), 140–145.
12. Latanska, L. O., & Polovynka, A. V. (2025). Matematychnе modeliuвання v zadachakh otsiniuvannya tryvalosti rozrobky bankivskoho prohramnoho zabezpechennia [Mathematical modeling in problems of estimating the duration of banking software development]. Proceedings of IX-th International scientific-technical conference Computer modeling and optimization of complex systems (CMOCS-2025) (p. 59). Dnipro, Ukraine. [in Ukrainian].
13. Prykhodko, N. V., & Prykhodko, S. B. (2018). Constructing nonlinear regression models on the basis of multivariate normalizing transformations. *Electronic Modeling*, 40(6), 99–108. <https://doi.org/10.15407/emodel.40.06.101>
14. Prykhodko, S. B., Prykhodko, N. V., Koltsov, A. V. (2024). A Nonlinear Regression Model for Early LOC Estimation of Open-Source Kotlin-Based Applications. *Radio Electronics Computer Science Control*, 1, 85–95. <https://doi.org/10.15588/1607-3274-2024-1-8>

Латанська Людмила Олексіївна – к.ф.-м.н., доцент, доцент кафедри програмного забезпечення автоматизованих систем Національного університету кораблебудування імені адмірала Макарова. E-mail: [liudmyla.latanska@nuos.edu.ua](mailto:liudmyla.latanska@nuos.edu.ua), ORCID: 0000-0001-6473-7624.

Половинка Антон Володимирович – аспірант кафедри програмного забезпечення автоматизованих систем Національного університету кораблебудування імені адмірала Макарова. E-mail: [anton.polovynka@nuos.edu.ua](mailto:anton.polovynka@nuos.edu.ua), ORCID: 0009-0007-8648-135X.

Latanska Liudmyla Oleksiivna – Candidate of Physical and Mathematical Sciences, Associate Professor, Associate Professor at the Department of Software for Automated Systems of the Admiral Makarov National University of Shipbuilding. E-mail: [liudmyla.latanska@nuos.edu.ua](mailto:liudmyla.latanska@nuos.edu.ua), ORCID: 0000-0001-6473-7624.

Polovynka Anton Volodymyrovych – Postgraduate Student at the Department of Software for Automated Systems of the Admiral Makarov National University of Shipbuilding. E-mail: [anton.polovynka@nuos.edu.ua](mailto:anton.polovynka@nuos.edu.ua), ORCID: 0009-0007-8648-135X.

Дата першого надходження статті до видання: 04.04.2026

Дата прийняття статті до друку після рецензування: 12.05.2026

Дата публікації (оприлюднення) статті: 01.07.2026



Стаття поширюється на умовах ліцензії відкритого доступу (CC BY 4.0)