

Державний комітет статистики України

ЗАТВЕРДЖЕНО

Наказ Держкомстату

01.10.2010 № 408

**зі змінами, затвердженими
наказом Державної служби
статистики України
від 28.12.2018 № 293**

МЕТОДОЛОГІЧНІ ПОЛОЖЕННЯ

**для оцінки показників бідності на регіональному рівні
із застосуванням методів статистико-математичного моделювання**

Методологічні положення розроблені з метою визначення загальних принципів застосування методів статистико-математичного моделювання для підвищення якості вимірювання основних показників бідності для регіонів України. У документі наведені прототипи моделей та підходи до визначення надійності оцінювання показників бідності на основі моделей.

Основну увагу приділено методологічним підходам до побудови моделей із використанням даних обстеження умов життя домогосподарств та зовнішньої інформації.

Розглянуті особливості побудови моделей для визначених показників бідності, надійність оцінювання яких необхідно підвищити на основі методів моделювання. Надано опис методів композиційного оцінювання з використанням прямих оціночних функцій, а також синтетичних оціночних функцій, які представлені моделями множинної лінійної регресії. Визначені основні вимоги до застосування статистико-математичних моделей оцінювання показників бідності та оцінки їх адекватності, а також методологічні підходи до створення форматів подання для користувачів результатів застосування методів моделювання.

Методологічні положення підготовлені експертом проекту Світового банку "Удосконалення системи соціальної допомоги в Україні" В.Г.Саріогло з урахуванням рекомендацій міжнародного експерта Т.Лонгфорда за участю фахівців департаменту обстежень домогосподарств Держкомстату:

Осипової І.І. – директора департаменту обстежень домогосподарств;

Спесивої–Ухової С.В. – заступника директора цього департаменту;

Пліско К.Ю. – заступника директора департаменту - начальника відділу методологічного та організаційного забезпечення обстеження умов життя домогосподарств.

Методологічні положення призначені для використання працівниками органів державної статистики, міністерств, інших органів виконавчої влади, науковими організаціями, навчальними закладами, іншими користувачами статистичної інформації.

Зміст

Визначення термінів	4
Передмова	6
1. Загальні положення	7
2. Методологічні підходи до оцінювання показників бідності на регіональному рівні на основі композиційних оціночних функцій	10
3. Джерела додаткової інформації для побудови синтетичних оціночних функцій ..	14
4. Характеристики надійності оцінювання показників бідності за моделями	17
5. Методологічні підходи до створення форматів подання для користувачів результатів застосування методів моделювання.....	18
Додатки	27
Додаток 1. Надійність прямих оцінок визначених показників бідності для націо- нального та регіонального рівня	27
Додаток 2. Опис узагальненого методу найменших квадратів.....	29
Додаток 3. Опис методу моментів.....	29
Додаток 4. Приклад побудови прототипів моделей.....	30
Додаток 5. Рівні бідності за національною межею та логарифми від них за регіонами України	38
Додаток 6. Кінцеві споживчі витрати домогосподарств у розрахунку на одну особу та логарифми від них за регіонами України.....	39
Додаток 7. Проект методологічних пояснень щодо оцінювання показників бідності з використанням методів математико-статистичного моделювання.....	40
Перелік використаних джерел	46

Визначення термінів

<i>Дизайн (план) вибірки</i>	– організаційно-логічна модель структури вибіркової сукупності та принципи її формування. Якістю плану вибірки в значній мірі визначається репрезентативність вибіркової сукупності, отриманої за визначеними у плані правилами
<i>Дисперсія вибірових оцінок</i>	– середній квадрат відхилень оцінок показника за всіма можливими вибірками даного дизайну (реплікаціями) від середнього значення оцінки за всіма вибірками
<i>Додаткова (зовнішня) інформація</i>	– у контексті даних методологічних положень це інформація, яка використовується при оцінюванні показників за результатами проведеного вибіркового обстеження, але отримана з іншого джерела даних (наприклад, це дані інших статистичних обстежень, дані цього ж обстеження, але по інших територіях або за інші періоди часу, дані статистичної звітності, адміністративні дані тощо)
<i>Коефіцієнт варіації (відносна стандартна похибка)</i>	– відношення величини стандартної похибки вибірки до оцінки показника (частіше за все вимірюється у відсотках)
<i>Композиційна оціночна функція</i>	– оціночна функція, отримана на основі композиції (комбінації) різних оціночних функцій. В офіційній статистиці частіше за все використовуються композиція прямих та непрямих оціночних функцій
<i>Непряме оцінювання</i>	– оцінювання показника на основі непрямой оціночної функції, при побудові якої передбачається використання даних, отриманих не лише за результатами безпосереднього вимірювання (на основі конкретного обстеження), а й з використанням інформації з додаткових джерел (наприклад, даних аналогічних обстежень по інших територіях, за інші періоди часу, дані інших обстежень тощо)
<i>Оцінка показника</i>	– значення показника, яке розраховано на основі відповідної оціночної функції за результатами вибіркового обстеження
<i>Оціночна функція</i>	– правило або співвідношення (формула) для оцінювання показника з використанням даних вибіркового обстеження. Оціночна функція відома ще до отримання результатів обстеження. Слід розрізняти поняття <i>оціночна функція</i> та <i>оцінка</i> , так як оціночна функція – це функція вибірки, тоді як оцінка є її значенням, знайденим на основі реалізації вибірки
<i>Оцінювання для малих територій</i>	– статистична методологія, яка розроблена з метою підвищення надійності оцінювання показників за результатами статистичних спостережень на основі прямих та непрямих оціночних функцій, комплексного використання інформації з різних джерел та сучасних методів статистико-математичного моделювання
<i>Пряме оцінювання</i>	– оцінювання показника на основі прямої оціночної функції, при побудові якої передбачається, що оцінювання здійснюється на основі даних, отриманих за результатами безпосереднього вимірювання (на основі конкретного обстеження)

<i>Середньоквадратична похибка</i>	– загальна похибка оцінювання показника, яка визначається як сума дисперсії вибірових оцінок показника та квадрата величини зміщення оцінки, яке є різницею між очікуваною величиною оцінки (тобто середнім із вибірових оцінок) та дійсним значенням показника для генеральної сукупності
<i>Стандартна похибка</i>	– середньоквадратичне відхилення вибірових оцінок показника за всіма можливими вибірками даного дизайну від дійсного значення показника для генеральної сукупності
<i>Статистико-математичне моделювання</i>	– методи математичної статистики, що забезпечують вибір серед можливих теоретико-ймовірнісних моделей такої, яка найкращим чином (у певному визначеному сенсі) відповідає наявним статистичним даним, що характеризують досліджуване явище або процес

Передмова

Реалізація ефективної соціальної політики в Україні потребує проведення постійного моніторингу бідності на загальнодержавному та регіональному рівнях, що вимагає надійного вимірювання основних показників бідності. Достатній рівень надійності зазначених показників забезпечує адекватність аналізу бідності, в т.ч. оцінки впливу на бідність заходів соціальної політики та, певною мірою, є базою для прийняття дієвих управлінських рішень у цій сфері.

В Україні показники бідності вимірюються за результатами державного вибіркового обстеження умов життя домогосподарств (далі – ОУЖД), методологія проведення якого відповідає міжнародним стандартам для обстежень такого типу.

За сучасними міжнародними рекомендаціями для державних вибірових обстежень населення (домогосподарств) надійність оцінок показників для країни в цілому, яка характеризується коефіцієнтом варіації CV , повинна забезпечуватися на рівні $CV \leq 5\% \div 8\%$. Для окремих територій (регіонів, районів) оцінки показників вважаються достатньо надійними, якщо коефіцієнт варіації не перевищує 10%, а для деяких обстежень і 15%.

Як свідчать результати проведених досліджень, в Україні для вимірювання бідності зазначені вимоги щодо надійності оцінювання показників бідності задовольняються не повною мірою. При існуючій системі показників бідності та наявній інформаційній базі статистична надійність показників, визначених на рівні регіонів, є недостатньою. Це призводить до нестабільності оцінок у динаміці та обмежує можливість адекватного моніторингу бідності.

Ефективним способом підвищення надійності показників бідності для різних рівнів агрегації даних є застосування спеціальних методологічних підходів, що базуються на методах непрямого оцінювання. Слід зазначити, що можливість впровадження таких методів, як правило, вимагає певного ускладнення технології обробки даних обстеження, відповідного удосконалення організаційних структур та підвищення рівня кваліфікації фахівців, які займаються розрахунками показників бідності.

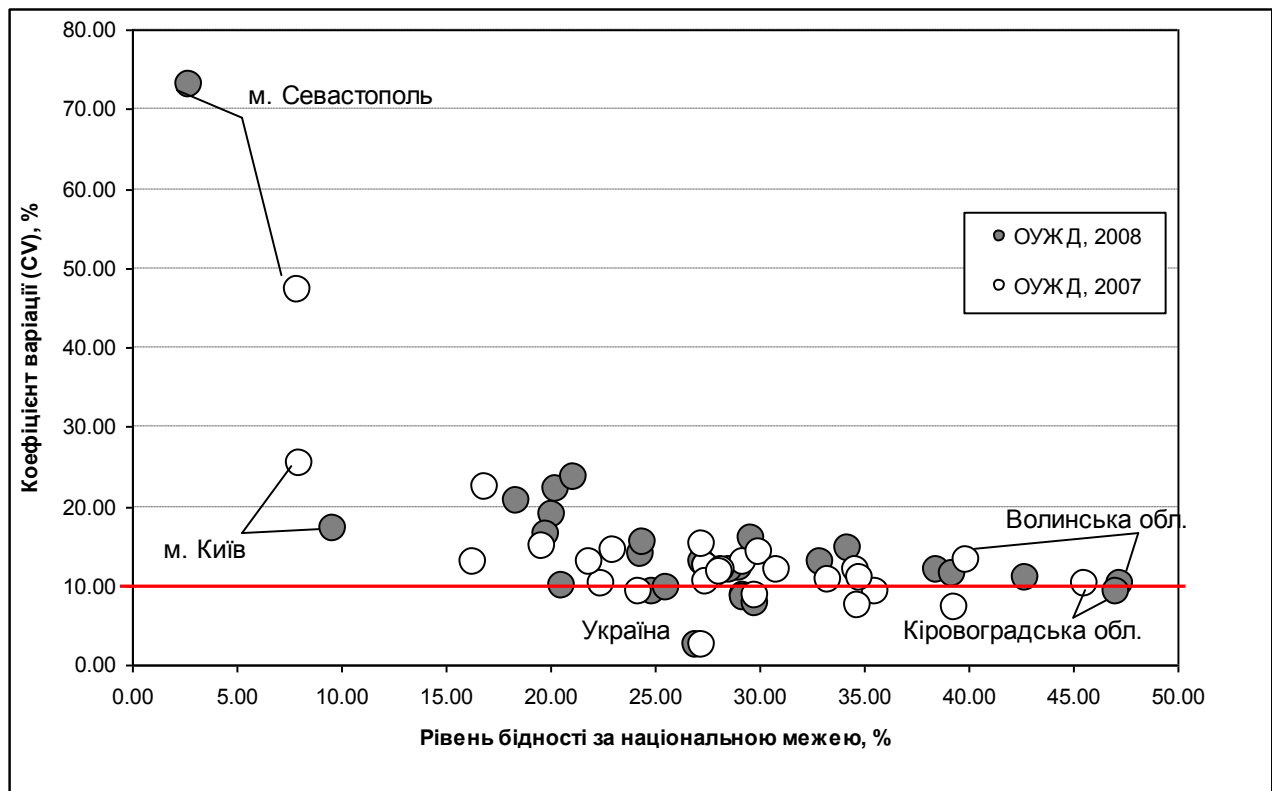
На основі результатів опрацювання існуючих підходів до підвищення рівня надійності результатів вибірових обстежень домогосподарств слід зробити висновок, що в теперішній час найбільш ефективною є методологія уточнення оцінок показників на базі так званих методів оцінювання для малих територій. Специфічна особливість цих методів полягає в комплексному використанні існуючої інформаційної бази та існуючих додаткових даних.

1. Загальні положення

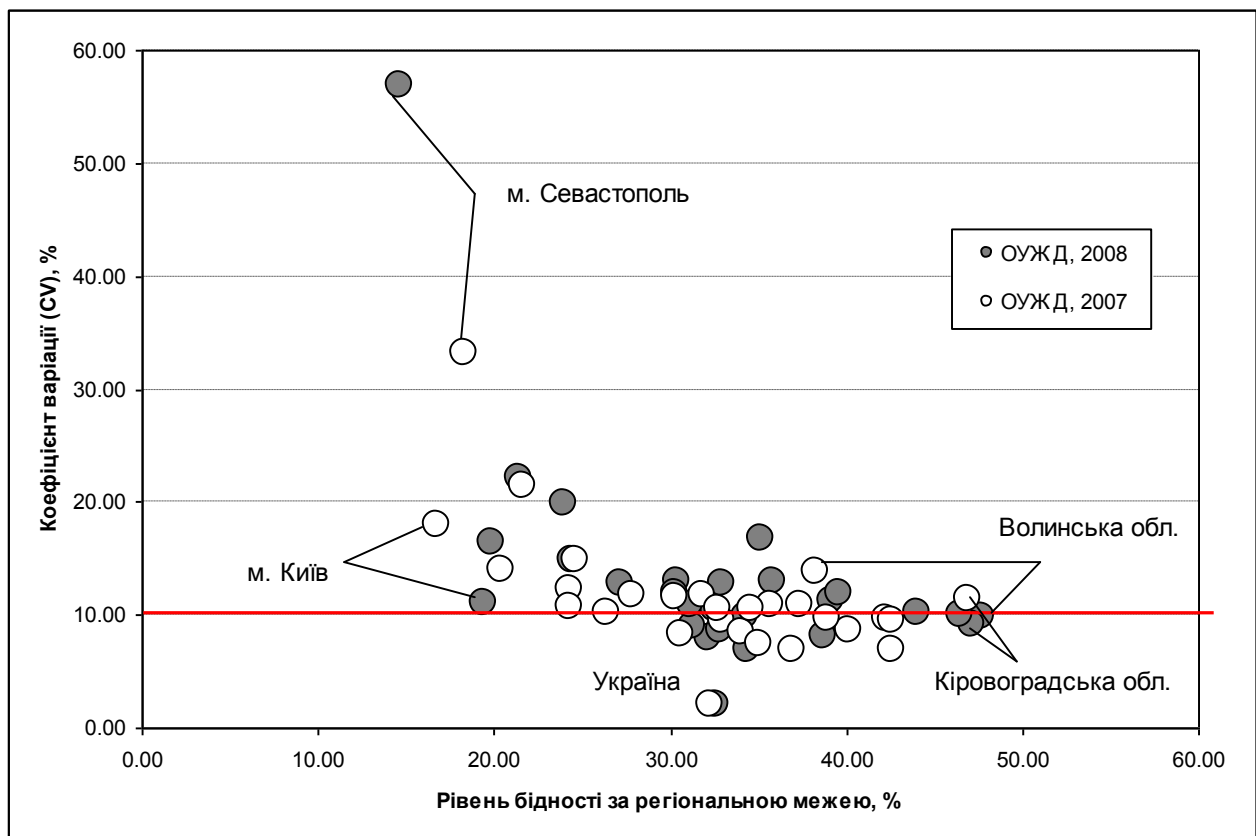
Проблема недостатньої надійності прямого оцінювання показників для окремих територій або груп одиниць генеральної сукупності за результатами вибірових обстежень відома як проблема малих територій. Недостатній рівень надійності прямих оцінок основних показників бідності для регіонів України є прикладом проблеми малих територій.

Пряме оцінювання показників може бути здійснене для національного рівня або для регіонів за стандартними прямими оціночними функціями [1–7]. За такого підходу оцінка показника для регіону заснована виключно на даних ОУЖД по цьому регіону. У багатьох випадках цього виявляється недостатньо для надійного оцінювання, що пояснюється відносно малими обсягами вибірки, на основі яких здійснюється оцінювання показників для регіонів, у порівнянні з обсягом вибірки для оцінювання показників на національному рівні.

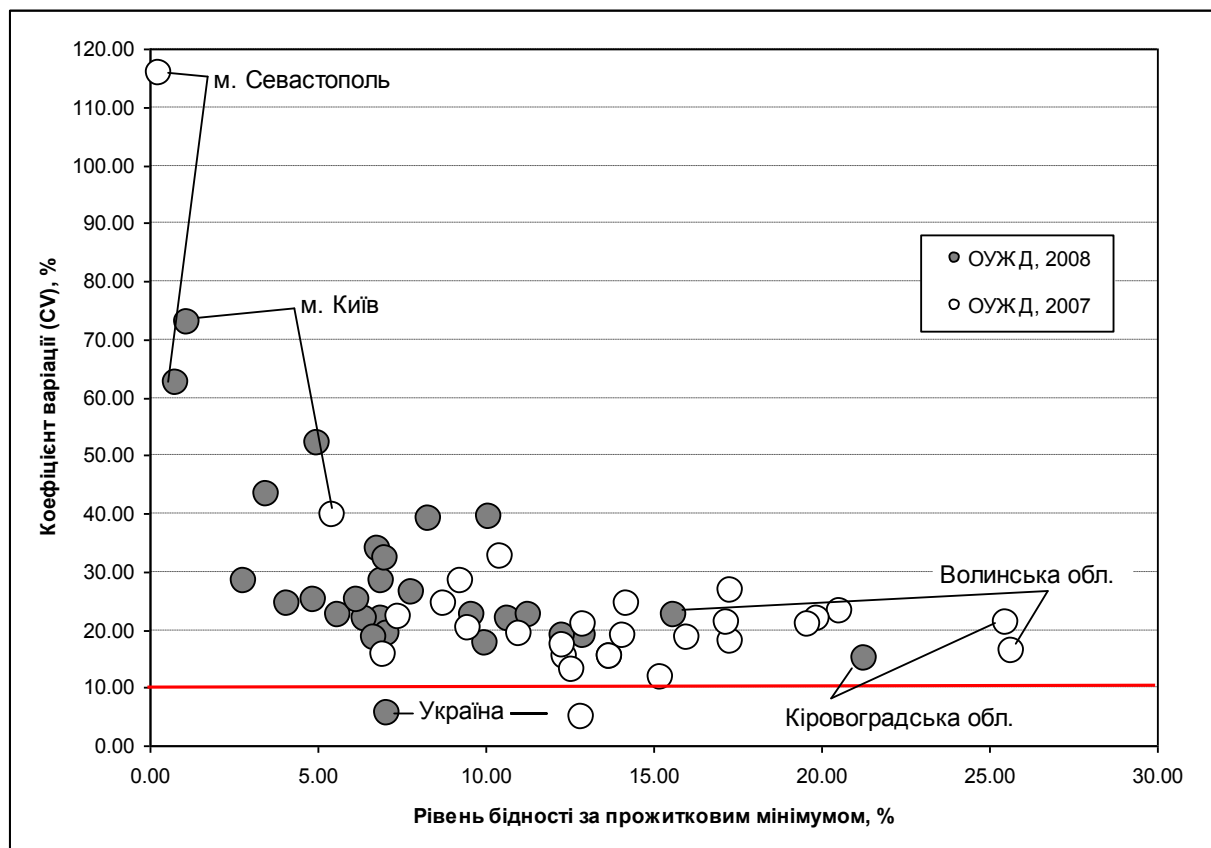
Спільним рішенням експертів, які займаються питаннями моніторингу бідності в Україні, були визначені основні показники бідності, надійність оцінювання яких має бути забезпечена на належному рівні (рівень бідності за національною межею, рівень бідності за регіональною межею та рівень бідності за прожитковим мінімумом [8–10]). Результати розрахунків засвідчують, що для цих визначених основних показників бідності надійність прямих оцінок на регіональному рівні є незадовільною для багатьох регіонів (див. рис. 1, а–в) та додаток 1).



а)



б)



в)

Рис.1. Надійність прямих оцінок визначених показників бідності для національного та регіонального рівня (ОУЖД, 2008): а) – рівень бідності за національною межею; б) – рівень бідності за регіональною межею; в) – рівень бідності за прожитковим мінімумом

На сучасному етапі підвищення надійності оцінювання показників за результатами вибірових обстежень може бути досягнуто на основі застосування спеціальної методології оцінювання – так званих методів малих територій. Головною ідеєю методів оцінювання для малих територій є те, що для підвищення надійності оцінювання показників крім безпосередніх результатів певного обстеження використовується інформація з додаткових джерел даних на рівні малої території, мікрорівні або на рівні більшої території, до складу якої входить мала територія, що розглядається. Наприклад, це можуть бути дані попередніх етапів обстеження, кореляти показників бідності, отримані з адміністративних джерел, за даними статистичної звітності, за результатами переписів тощо.

Серед традиційних методів малих територій, які пропонуються для застосування в офіційній статистиці для вирішення проблеми недостатньої надійності оцінок показників, пріоритетне значення має метод композиційного оцінювання (composite estimation) [3–7]. Історично метод композиційного оцінювання виник у результаті пошуку можливостей об'єднання (композиції) прямих і непрямих оціночних функцій. При цьому в якості непрямих можуть використовуватись так звані синтетичні оціночні функції, побудовані на основі статистичних моделей із використанням додаткової інформації [4]. В сучасній статистичній практиці цей метод став ефективним інструментом розв'язання проблеми підвищення надійності оцінок показників на різних рівнях агрегації даних.

При композиційному оцінюванні певного цільового показника Y для регіону k оціночна функція може бути представлена в наступному спрощеному вигляді:

$$\hat{Y}_k^{(H)} = \gamma_k \cdot \hat{Y}_k^{(D)} + (1 - \gamma_k) \cdot \hat{Y}_k^{(S)}, \quad (1)$$

де $\hat{Y}_k^{(H)}$ – композиційна оціночна функція; $\hat{Y}_k^{(D)}$ – пряма оціночна функція; $\hat{Y}_k^{(S)}$ – синтетична оціночна функція; γ_k – ваговий коефіцієнт, який, як правило, визначається з умови мінімізації середньоквадратичної похибки оцінювання (цей коефіцієнт приймає значення від 0 до 1); $k = 1, 2, \dots, K$, K – загальна кількість регіонів.

Слід зауважити, що специфічною особливістю даного методу є те, що композиційна оціночна функція дозволяє отримувати оцінки показників навіть для територій, які не увійшли до територіальної вибірки і по яких, відповідно, неможливо отримати прямі оцінки. У таких випадках значення коефіцієнта γ_k у формулі (1) дорівнює 0.

При практичному впровадженні методу композиційного оцінювання слід враховувати, що його застосування доцільне для уточнення лише певних найбільш важливих показників, що обумовлено складністю визначення зовнішньої інформації та побудови непрямих оціночних функцій. Крім того, необхідно враховувати такі загальні вимоги до цього методу оцінювання

показників бідності, встановлені експертами: розроблені статистико-математичні моделі повинні забезпечувати оцінювання визначених показників бідності для всіх регіонів України з надійністю $CV \leq 10\%$ на основі річного масиву даних ОУЖД та визначеної зовнішньої інформації (виключенням є м. Київ та м. Севастополь, можливість забезпечення припустимого рівня надійності показників бідності по яких потребує окремого дослідження, що пояснюється відносно незначними рівнями бідності населення в цих регіонах); джерела інформації, що використовуватиметься для уточнення показників бідності на основі моделей, повинні бути доступними на регулярній основі; розроблені моделі та алгоритми повинні бути зручними для автоматизації програмними засобами, що використовуються в системі державної статистики України. Терміни обробки даних для моделювання та розрахунків уточнених показників повинні бути мінімізовані наскільки це можливо.

2. Методологічні підходи до оцінювання показників бідності на регіональному рівні на основі композиційних оціночних функцій

Для підвищення надійності оцінювання визначених показників бідності на регіональному рівні в Україні передбачається використання композиційних оціночних функцій. Для адекватного визначення методологічних основ побудови цих оціночних функцій доцільно розглянути ряд базових показників та їх позначень.

Нехай за результатами ОУЖД за поточний рік розраховано оцінку певного з визначених основних показників бідності для k -го регіону $\hat{p}_k$ за прямою оціночною функцією (наприклад, це може бути оцінка рівня бідності за національною межею). Для узагальнення через $\hat{\theta}_k$ позначимо пряму оціночну функцію $\hat{p}_k$ або певну математичну функцію від неї, наприклад, логарифм від $\hat{p}_k$. Припустимо, що додаткова (зовнішня) інформація наявна для регіону k у вигляді деяких показників $x_{1,k}, x_{2,k}, \dots, x_{J,k}$, де J – загальна кількість показників із додаткових джерел інформації. Нехай $\hat{\theta}_k^{(s)}$ – це позначення узагальненої синтетичної оціночної функції. Як свідчать результати міжнародних досліджень, для оцінювання показників бідності ефективним є представлення синтетичної оціночної функції моделлю лінійної множинної регресії [4, 5, 7]:

$$\hat{\theta}_k^{(s)} = x_k^T \hat{\beta}, \quad (2)$$

де x_k^T – вектор-рядок зовнішніх змінних:

$$x_k^T = \{1, x_{1,k}, x_{2,k}, \dots, x_{J,k}\}, \quad (3)$$

$\hat{\beta}$ – вектор-стовпець коефіцієнтів регресії:

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \dots \\ \hat{\beta}_J \end{bmatrix}. \quad (4)$$

Таким чином, синтетична оціночна функція для регіону k може бути представлена у вигляді:

$$\hat{\theta}_k^{(S)} = \mathbf{x}_k^T \hat{\beta} = \{1, x_{1k}, x_{2k}, \dots, x_{Jk}\} \cdot \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \dots \\ \hat{\beta}_J \end{bmatrix} = \hat{\beta}_0 + \hat{\beta}_1 x_{1k} + \hat{\beta}_2 x_{2k} + \dots + x_{Jk} \hat{\beta}_J. \quad (5)$$

Необхідно зазначити, що за підходом, який розглядається, регресійна модель є загальною для всіх регіонів, тобто коефіцієнти регресії не залежать від регіону k .

За прийнятих умов композиційна оціночна функція для k -го регіону $\hat{\theta}_k^{(H)}$ може бути представлена як середньозважена прямої оціночної функції $\hat{\theta}_k$ та синтетичної оціночної функції $\hat{\theta}_k^{(S)}$ у вигляді:

$$\hat{\theta}_k^{(H)} = \gamma_k \hat{\theta}_k + (1 - \gamma_k) \hat{\theta}_k^{(S)}, \quad (6)$$

де γ_k – ваговий коефіцієнт композиційної оціночної функції, $0 \leq \gamma_k \leq 1$.

Ваговий коефіцієнт γ_k визначається з умови мінімізації середньоквадратичної похибки оціночної функції $\hat{\theta}_k^{(H)}$ на основі формули [5, 7]:

$$\gamma_k = \frac{\sigma_v^2}{(\psi_k + \sigma_v^2)}, \quad (7)$$

де σ_v^2 – дисперсія синтетичної оціночної функції (похибка за моделлю множинної лінійної регресії); ψ_k – дисперсія вибірових оцінок прямої оціночної функції $\hat{\theta}_k$.

Коефіцієнти регресії синтетичної оціночної функції визначаються з використанням узагальненого методу найменших квадратів (див. додаток 2). Слід зазначити, що коефіцієнти регресії залежать від величини дисперсії синтетичної оціночної функції σ_v^2 .

Виходячи з формул (6) та (7) при оцінюванні показників бідності за методом, що розглядається, більшу вагу має та оціночна функція, надійність якої є вищою. За умови, що для k -го регіону похибка за моделлю множинної лінійної регресії σ_v^2 дорівнює дисперсії вибіркового оцінок прямої оціночної функції ψ_k , величина вагового коефіцієнту композиційної оціночної функції складатиме 0,5.

Середньоквадратична похибка оціночної функції (6) може бути наближено представлена у вигляді [4]:

$$MSE(\hat{\theta}_k^{(H)}) = \gamma_k \psi_k + (1 - \gamma_k)^2 x_k^T \left[\sum_{k=1}^K \frac{x_k x_k^T}{\psi_k + \sigma_v^2} \right]^{-1} x_k. \quad (8)$$

Величина середньоквадратичної похибки композиційної оцінки $MSE(\hat{\theta}_k^{(H)})$, як і коефіцієнти регресії, залежить від величини дисперсії синтетичної оціночної функції σ_v^2 .

Для оцінки величини σ_v^2 , за наявності даних обстеження та зовнішньої інформації, можуть застосовуватись метод моментів (див. додаток 3), метод максимальної правдоподібності (ML – Maximum Likelihood) або максимальної правдоподібності з обмеженнями ($REML$ – Restricted Maximum Likelihood) [4, 5, 7].

Всі зазначені процедури оцінювання σ_v^2 , як і процедури розрахунку коефіцієнтів регресії синтетичної оціночної функції, детально розглянуті в існуючих публікаціях з питань оцінювання показників для малих територій [4, 5, 7] і реалізовані у статистичному програмному забезпеченні, зокрема в R , яке працює як незалежно, так і у середовищі $SPSS$.

Таким чином, статистична ефективність оцінювання показників бідності за композиційною оціночною функцією (6) суттєво залежить від якості синтетичної оціночної функції (моделі множинної лінійної регресії), що, в свою чергу, вимагає визначення набору корелят для кожного показника бідності, надійність оцінювання якого необхідно підвищити.

В цілому побудова композиційної оціночної функції для уточнення оцінок визначених показників бідності за регіонами складається з таких основних етапів:

- визначення прямих оцінок показників бідності за регіонами;
- оцінка дисперсій вибіркового оцінок для прямих оцінок показників бідності $\hat{\psi}_k$ за регіонами;

- підбір зовнішньої інформації та побудова наближеної моделі регресії для визначення синтетичної оціночної функції. Параметри наближеної моделі регресії визначаються за стандартними процедурами регресійного аналізу без урахування похибки моделі регресії $\hat{\sigma}_v^2$ та дисперсій вибірових оцінок для прямих оцінок показників бідності $\hat{\psi}_k$;
- оцінка похибки моделі регресії $\hat{\sigma}_v^2$ та коефіцієнтів регресії синтетичної оціночної функції за алгоритмами, які відповідають методу композиційного оцінювання (див. додатки 2 і 3);
- оцінка вагових коефіцієнтів композиційної оціночної функції $\hat{\gamma}_k$;
- розрахунки композиційних оцінок показників бідності для регіонів.

Враховуючи, що при використанні запропонованих композиційних оціночних функцій існує проблема побудови та використання синтетичних оціночних функцій, які представляються моделями лінійної множинної регресії, доцільно навести критерії, що використовуються для вибору найбільш ефективних рівнянь регресії.

Для оцінки якості моделей використовуються стандартні процедури [8, 11, 12]. Основними характеристиками адекватності моделі є коефіцієнт детермінації R^2 та F -відношення.

Коефіцієнт детермінації R^2 змінюється від 0 до 1 і показує частку змін результативної ознаки (дисперсії результативної ознаки), яка пояснюється регресійною моделлю. Чим ближчі значення R^2 до 1, тим краще лінійна регресійна модель апроксимує результативну ознаку. У загальному випадку критичними для адекватності регресійної моделі слід вважати значення R^2 , які є меншими за 0,7. F -відношення відображає ступінь взаємозв'язку між результативною і факторними ознаками. Якщо розрахункове значення F -відношення перевищує порогове значення $F_{кр}$ (порогове значення визначається за спеціальними формулами або з довідників, див., наприклад, [10]), то це свідчить про наявність взаємозв'язку між результативною та факторними ознаками. Таким чином, про адекватність моделі регресії наближено можна зробити висновок, якщо $R^2 \geq 0,7$ та $F > F_{кр}$. Але слід враховувати, що моделі регресії мають багато інших параметрів, які можуть бути проаналізовані для обґрунтування моделі або покращення її адекватності, це і аналіз залишків, і аналіз похибок визначення коефіцієнтів регресії тощо [11, 12].

Слід зазначити, що кінцевим критерієм ефективності оціночної функції є величина $MSE(\hat{\theta}_k^{(H)})$. Тому кінцеву синтетичну оціночну функцію необхідно обирати саме за найменшою (значимо) величиною $MSE(\hat{\theta}_k^{(H)})$ композиційної оціночної функції.

При реалізації на практиці підходів до оцінювання показників на основі моделей необхідно приділяти особливу увагу всебічному аналізу ефективності моделей регресії, що

застосовуються. З можливих варіантів для практичного використання доцільно відбрати найбільш прості моделі, що забезпечує більшу ясність процедур при задовільному рівні точності оцінювання.

Доцільно враховувати, що композиційне оцінювання показників бідності за регіонами вимагає перерахунку відповідних показників рівня України та, наприклад, рівня економічних районів. При цьому для розрахунку показника на національному рівні застосовується така загальна формула:

$$\hat{\theta}_U^{(H)} = \frac{\sum_{k=1}^K N_k \hat{\theta}_k^{(H)}}{\sum_{k=1}^K N_k}, \quad (9)$$

де $\hat{\theta}_U^{(H)}$ – уточнена оцінка показника бідності на національному рівні; $\hat{\theta}_k^{(H)}$ – композиційна оцінка показника бідності для k -го регіону; N_k – середньорічна чисельність населення k -го регіону.

Для економічних районів уточнені оцінки показників бідності розраховуються за аналогічними формулами з урахуванням даних лише по тих регіонах, які входять до складу відповідних економічних районів.

Середньоквадратичні похибки оцінок показників бідності на національному рівні $MSE(\hat{\theta}_U^{(H)})$ визначаються на основі середньоквадратичних похибок оцінок показників для регіонів за формулою:

$$MSE(\hat{\theta}_U^{(H)}) = \frac{1}{N^2} \sum_{k=1}^K N_k^2 MSE(\hat{\theta}_k^{(H)}). \quad (10)$$

Приклад реалізації запропонованих методологічних підходів з метою підвищення надійності оцінювання рівня бідності за національною межею наведено у додатку 4.

3. Джерела додаткової інформації для побудови синтетичних оціночних функцій

Джерела, структура та обсяги інформації, яка може бути використана для цілей підвищення надійності оцінювання показників бідності, визначаються діючою методологією вимірювання бідності, а також наявністю та доступністю надійних статистичних та адміністративних даних (зовнішньої інформації), які характеризують рівень життя населення [9].

При відборі джерел зовнішніх даних для оцінювання показників бідності із застосуванням моделей слід брати до уваги, що використання значної кількості джерел призводить, як правило, до зменшення величини зміщення оцінок за моделями і забезпечує повноту врахування всіх можливих взаємозв'язків показників бідності із зовнішніми даними, але майже завжди обумовлює зростання величини дисперсії оцінок показників, що пояснюється зростанням нестабільності оцінок параметрів моделей [5, 6]. Головною особливістю зовнішньої інформації, яка може бути використана при оцінюванні показників бідності із застосуванням моделей, є те, що вона повинна бути статистично тісно пов'язаною з відповідними результатами ОУЖД і бути достатньо надійною. Це, певною мірою, ускладнює можливість отримання інформації, корисної для оцінювання показників, з інших джерел. Відповідно, перелік джерел зовнішньої інформації, які доцільно використовувати з метою підвищення надійності оцінювання показників бідності, частіше за все є достатньо обмеженим [6].

Таким чином, інформаційною базою для підвищення надійності оцінювання показників бідності для регіонів України на основі статистико-математичних моделей можуть бути, насамперед, ті джерела, які дають можливість оцінити показники, що статистично тісно пов'язані з показниками рівня життя населення. Такими джерелами інформації є, наприклад, державні вибіркові обстеження домогосподарств (населення) і, перш за все, безпосередньо дані ОУЖД [10]. Іншими потенційними джерелами є дані статистики праці, соціальної статистики, національних рахунків, переписів населення, адміністративні дані.

Цінність даних ОУЖД обумовлена їх методологічною узгодженістю, що забезпечує порівнянність показників за різні періоди часу та по різних територіях. Як свідчать результати виконаних досліджень, для уточнення показників бідності на даному етапі розробки та впровадження статистико-математичних моделей для надійного оцінювання показників бідності, насамперед доцільно розглядати прямі оцінки визначених показників бідності за попередні періоди обстеження; прямі оцінки показників бідності за результатами поточного обстеження для територій, що включають регіони, наприклад, це можуть бути оцінки для національного рівня та рівня економічних районів.

Водночас слід зазначити, що використання лише показників ОУЖД для уточнення показників бідності характеризується і певними недоліками. Основний недолік полягає у тому, що у випадку зміщеності оцінок показників ОУЖД (наприклад, внаслідок недостатнього охоплення заможних верств населення), непрямі оцінки показників бідності будуть також зміщеними і, відповідно, ступінь реального підвищення надійності їх оцінювання буде в цілому незначним. Інший суттєвий недолік полягає в тому, що оцінки показників бідності та оцінки інших показників ОУЖД не є статистично незалежними, що обумовлено використанням фактично однієї і тієї ж інформаційної бази для розрахунків. Таким чином, доцільно використовувати поряд із наведеними даними також дані, отримані за результатами вибірових

обстежень, не пов'язаних з ОУЖД, даних статистичної звітності, даних адміністративних реєстрів тощо.

Як свідчать результати аналізу щільності взаємозв'язку показників державного вибіркового обстеження робочої сили (далі – ОРС) та показників бідності, вона є недостатньо високою. Так, для визначених показників бідності та рівня безробіття, за даними ОРС, коефіцієнт кореляції не перевищує 0,5. Коефіцієнт кореляції зазначених показників бідності з рівнем зайнятості складає від $-0,3$ до $-0,4$. Виходячи з отриманих результатів, слід констатувати, що використання показників ОРС в якості зовнішніх змінних при розробці моделей для уточнення показників бідності є недоцільним.

Коефіцієнт кореляції показників бідності з рівнем зареєстрованого безробіття складає приблизно від 0,3 до 0,5, що в цілому не свідчить про наявність щільного взаємозв'язку. Оскільки рівень зареєстрованого безробіття є адміністративною інформацією, то статистична надійність даних приймається максимальною, тобто дані вважаються абсолютно надійними в тому сенсі, що випадкова похибка розрахунку показників приймається рівною нулю. Систематичні похибки (можливі зміщення значень показників), пов'язані з методологією вимірювання показників зареєстрованого безробіття, похибками обробки даних тощо, не приймаються до уваги, оскільки вони не впливають на коефіцієнти кореляції, що характеризують рівень взаємозв'язку між розподілами показників. Це зауваження стосується всіх джерел адміністративної інформації або статистичної інформації, що отримана на основі статистичної звітності.

Як свідчать результати досліджень, при непрямому оцінюванні окремих із визначених показників бідності, корисним може бути використання розподілу частки працівників, що отримують зарплату нижче мінімального рівня (за даними статистики праці) за регіонами, середнього розміру призначених пенсій з цільовою допомогою за регіонами (за даними соціальної статистики), величини кінцевих споживчих витрат домогосподарств у розрахунку на одну особу за регіонами (за даними національних рахунків) [10].

Для вибору незалежних змінних при побудові моделі регресії (2), насамперед, аналізують парні коефіцієнти кореляції Пірсона незалежної змінної (прямих оцінок показників бідності або певних функцій від них $\hat{\theta}_k$) з кожною з L потенційних незалежних змінних $x_{1,k}, x_{2,k}, \dots, x_{L,k}$. Як правило, з сукупності потенційних незалежних змінних обираються лише ті змінні, коефіцієнти парної кореляції з якими перевищують 0,7, але в окремих випадках, особливо на пошуковому етапі, доцільно включити до аналізу також змінні з коефіцієнтами кореляції 0,5 і вище. При цьому застосовують також математичні перетворення змінних, наприклад, операції логарифмування, що може покращити статистичний взаємозв'язок розподілів, що аналізуються. Далі зі сформованої сукупності зовнішніх змінних із

використанням покрокової процедури (покрокової регресії) [11] формується оптимальний набір зовнішніх змінних $x_{1,k}, x_{2,k}, \dots, x_{J,k}$. Доцільно враховувати, що кількість незалежних змінних, яку можна використати у регресійній моделі, є обмеженою. Наприклад, якщо мова йде про 27 регіонів України, то в моделі може бути використано, як правило, не більше трьох незалежних змінних, що пояснюється особливостями оцінювання її параметрів у регресійному аналізі [11].

4. Характеристики надійності оцінювання показників бідності за моделями

Після визначення прототипу моделі для оцінювання показників бідності за основний критерій надійності доцільно прийняти умову, за якої значення відносної похибки оцінювання показника бідності повинно бути меншим за встановлене граничне значення цієї похибки. Такою відносною похибкою є величина відношення кореня квадратного від середньоквадратичної похибки до композиційної оцінки показника бідності за регіонами:

$$RRMSE(\hat{\theta}_k^{(H)}) = \frac{\sqrt{MSE(\hat{\theta}_k^{(H)})}}{\hat{\theta}_k^{(H)}} \cdot 100\%, \quad (11)$$

де середньоквадратична похибка $MSE(\hat{\theta}_k^{(H)})$ визначається за формулою (8).

Величина $RRMSE(\hat{\theta}_k^{(H)})$ є аналогом коефіцієнта варіації для прямих оцінок і відрізняється від цього показника тим, що враховує зміщення оцінок за моделлю. Межі значень $RRMSE(\hat{\theta}_k^{(H)})$, що визначають рівень надійності показників, такі ж, як і для коефіцієнтів варіації. Композиційна оцінка показника бідності може вважатися достатньо надійною для регіонального рівня, якщо значення $RRMSE(\hat{\theta}_k^{(H)})$ не перевищує 10%, в окремих випадках і 15%. Якщо значення $RRMSE(\hat{\theta}_k^{(H)})$ знаходиться в інтервалі 15–25%, використання показника рекомендується лише для якісного аналізу. Якщо значення $RRMSE(\hat{\theta}_k^{(H)})$ перевищує 25%, показник не рекомендується використовувати для аналізу бідності.

Для визначення та аналізу надійності композиційного оцінювання показників також використовують величину кореня квадратного від середньоквадратичної похибки $RMSE(\hat{\theta}_k^{(H)}) = \sqrt{MSE(\hat{\theta}_k^{(H)})}$, яка є аналогом стандартної похибки та величину граничної похибки вибірки $ME(\hat{\theta}_k^{(H)})$, яка, для довірчої ймовірності 95%, визначається за формулою:

$$ME(\hat{\theta}_k^{(H)}) = 1.96 \cdot RMSE(\hat{\theta}_k^{(H)}). \quad (12)$$

Для оцінки результатів моделювання здійснюються такі основні перевірки [13]:

- перевіряються значення оцінок вагових коефіцієнтів композиційної оціночної функції з метою підтвердження того, що значення потрапляють до визначеного інтервалу: $0 \leq \hat{\gamma}_k \leq 1$;
- перевіряються оцінки відносних похибок – відношень кореня квадратного від середньоквадратичної похибки до композиційної оцінки показника бідності за регіонами $RRMSE(\hat{\theta}_k^{(H)})$ – з метою підтвердження того, що їх значення відповідають встановленим вимогам щодо рівня надійності оцінювання показників бідності за регіонами;
- порівнюються прямі $\hat{\theta}_k$ та синтетичні $\hat{\theta}_k^{(S)}$ оцінки показників з метою перевірки того, що синтетичні оцінки потрапляють до довірчих інтервалів, побудованих для відповідних прямих оцінок – $\hat{\theta}_k \pm ME(\hat{\theta}_k)$. Це дозволяє виявити випадки, коли синтетична оцінка для окремого регіону відхиляється від прямої оцінки на значну величину, порівняно з довірчим інтервалом, що може свідчити про неадекватність моделі для цього регіону та вимагати проведення спеціального аналізу і, можливо, побудови іншої моделі.

5. Методологічні підходи до створення форматів подання для користувачів результатів застосування методів моделювання

У цьому розділі наведені методологічні підходи до розробки проекту вихідної інформації за результатами оцінювання визначених показників бідності для регіонів України з використанням статистико-математичних моделей. Ця вихідна інформація призначена як для внутрішнього використання з метою оцінки поточних результатів моделювання, так і для використання при підготовці інформації для надання зовнішнім користувачам [14].

У статистичній практиці інших країн оприлюднення результатів застосування моделей для підвищення надійності оцінювання показників здійснюється, як правило, один раз на рік із суттєвим запізненням даних, яке може складати один-два роки від звітного року. Це пояснюється, з одного боку, складністю розрахунків показників за моделями та оцінювання характеристик якості цих показників. З іншого боку, можливість застосування моделей вимагає наявності інформації з різних джерел, яка надходить неодноразово, і до якої також застосовуються спеціальні процедури перетворення, узгодження тощо. Важливим є також той факт, що зі зміною показників та зовнішньої інформації параметри моделей можуть потребувати певного коригування.

Результати моделювання, як правило, мають надаватися користувачам у вигляді окремої публікації, до складу якої доцільно включити методологічні пояснення. Нижче наведено стислий опис змісту методологічних пояснень щодо оцінювання показників бідності з

використанням методів статистико-математичного моделювання, розроблений на основі аналізу існуючих закордонних публікацій.

- Загальний опис проблеми

Пояснюється необхідність підвищення рівня надійності оцінювання визначених показників бідності за результатами вибіркового обстеження.

- Система показників

Зазначається, які саме показники бідності розраховано, методика (формула) розрахунку кожного показника.

- Рівень оцінювання

Зазначається, для якого саме територіального рівня (по яких розрізах) проводяться розрахунки показників бідності.

- Методи оцінювання показників

Вказується, на основі якого типу моделей отримано оцінки показників бідності для кожного територіального рівня. Надається характеристика зовнішньої інформації, що використовується. Наводяться характеристики адекватності моделей.

- Підходи до оцінювання надійності показників бідності

Наводяться характеристики та критерії надійності, формули розрахунку основних характеристик надійності, наведених у публікації.

- Технічні зауваження

Зазначається програмне забезпечення, на основі якого отримано оцінки показників бідності.

- Користування результатами

Зазначається, на яке коло користувачів розраховані наведені в публікації результати оцінки показників бідності.

- Додаткова інформація

Наводиться контактна інформація (поштова адреса, електронна адреса, веб-сторінка), за якою можна звернутися для отримання додаткової інформації щодо методології оцінювання показників бідності, результати застосування якої наведені в публікації.

Основні терміни та поняття

Надається визначення основних понять і термінів, які використовуються в даній публікації.

Проект методологічних пояснень наведено у додатку 7.

Вихідну інформацію за результатами моделювання доцільно розділити на два типи: для внутрішнього використання та для надання зовнішнім користувачам.

Вихідна інформація для внутрішнього використання

За результатами застосування оціночної функції (див. формули (5)–(12)) для оцінювання показників бідності по кожному з показників, надійність яких має бути підвищена, виводиться така інформація (див. табл. 1–2):

- прямі оцінки показника за регіонами, $\hat{\theta}_k$;
- оцінки дисперсії прямих оцінок за регіонами, $\hat{\psi}_k = \hat{V}_k(\hat{\theta}_k)$ та коефіцієнтів варіації, $CV_k = (\sqrt{\hat{\psi}_k} / \hat{\theta}_k) \cdot 100\%$;
- синтетичні оцінки показника за регіонами, $\hat{\theta}_k^{(S)} = x_k^T \hat{\beta}$;
- вектор-стовпець оцінки коефіцієнтів регресії, $\hat{\beta}$;
- коефіцієнт детермінації, R^2 ;
- F -відношення (F -критерій);
- оцінка дисперсії синтетичної оціночної функції, $\hat{\sigma}_v^2$;
- оцінки вагових коефіцієнтів композиційної оціночної функції за регіонами, $\hat{\gamma}_k$;
- композиційні оцінки показника за регіонами, $\hat{\theta}_k^{(H)}$;
- оцінки величини середньоквадратичних похибок композиційних оціночних функцій за регіонами $MSE(\hat{\theta}_k^{(H)})$;
- величини відношення кореня квадратного від середньоквадратичної похибки до композиційної оцінки показника бідності за регіонами $RRMSE(\hat{\theta}_k^{(H)})$;
- оцінка величини граничної похибки вибірки $ME(\hat{\theta}_k^{(H)})$ за регіонами.

Інформація для внутрішнього користування необхідна для визначення ефективності оцінювання показників для регіонів на основі прийнятої композиційної оціночної функції.

Для оцінки результатів моделювання здійснюються перевірки, викладені наприкінці розділу 4 цього документа, на базі таких основних показників:

- оцінок вагових коефіцієнтів композиційної оціночної функції;
- оцінок відносних похибок;
- порівняння прямих $\hat{\theta}_k$ та синтетичних $\hat{\theta}_k^{(S)}$ оцінок показників.

Таблиця 1

Макет вихідної інформації за регіонами за результатами застосування статистико-математичних моделей для оцінювання показників бідності для внутрішнього використання

	Пряма оцінка показника, $\hat{\theta}_k$	Дисперсія прямої оцінки $\hat{\psi}_k$	Коефіцієнт варіації прямої оцінки, CV_k	Довірчий інтервал для прямої оцінки	Синтетична оцінка, $\hat{\theta}_k^{(S)}$	Ваговий коефіцієнт, $\hat{\gamma}_k$	Композиційна оцінка, $\hat{\theta}_k^{(H)}$	Середньоквадратична похибка композиційної оцінки, $MSE(\hat{\theta}_k^{(H)})$	Гранична похибка композиційної оцінки, $ME(\hat{\theta}_k^{(H)})$	Відносна похибка композиційної оцінки, $RRMSE(\hat{\theta}_k^{(H)})$
АР Крим										
Вінницька область										
Волинська область										
Дніпропетровська область										
Донецька область										
Житомирська область										
Закарпатська область										
Запорізька область										
Івано-Франківська область										
Київська область										
Кіровоградська область										
Луганська область										
Львівська область										
Миколаївська область										
Одеська область										
Полтавська область										
Рівненська область										
Сумська область										
Тернопільська область										
Харківська область										
Херсонська область										
Хмельницька область										
Черкаська область										
Чернівецька область										
Чернігівська область										
м. Київ										
м. Севастополь										

Макет вихідної інформації за результатами застосування статистико-математичних моделей для оцінювання показників бідності для внутрішнього використання

Назва показника	Оцінка показника
Коефіцієнт регресії (вільний член), $\hat{\beta}_0$	
Коефіцієнт регресії, $\hat{\beta}_1$	
Коефіцієнт регресії*, $\hat{\beta}_2$	
Коефіцієнт детермінації, R^2	
F -відношення	
Дисперсії синтетичної оціночної функції, $\hat{\sigma}_v^2$	

*Для визначеності тут прийнято, що синтетична оціночна функція побудована на основі двох зовнішніх змінних: $\hat{\theta}_k^{(S)} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$.

Інші параметри моделей, що надаються внутрішнім користувачам, мають значення допоміжної інформації, яка може використовуватись для поглибленого аналізу результатів моделювання та контролю адекватності моделей. Наприклад, порівняння вихідної інформації поточного року з даними попереднього року може дозволити пояснити зміни окремих параметрів моделі (моделей) або рівнів надійності оцінювання показників. Про адекватність моделі регресії наближено можна зробити висновок якщо $R^2 \geq 0,7$ та $F > F_{кр}$. Якщо ці умови (або одна з умов) не виконуються, це може свідчити про неадекватність регресійної моделі і необхідність побудови іншої моделі. При виконанні зазначених умов, але суттєвих змінах коефіцієнтів регресії, доцільно проаналізувати кореляційні зв'язки залежних та незалежної змінних та з'ясувати причини змін. Величину $\hat{\sigma}_v^2$ доцільно порівняти з дисперсіями прямих оцінок $\hat{\psi}_k = \hat{V}_k(\hat{\theta}_k)$ – вона повинна бути близькою до середнього значення цих дисперсій або меншою за нього. Слід зазначити, що кінцевим критерієм ефективності оціночної функції є величина $MSE(\hat{\theta}_k^{(H)})$. Тому синтетичну оціночну функцію необхідно обирати саме за найменшою (значимо) величиною $MSE(\hat{\theta}_k^{(H)})$ композиційної оціночної функції.

Вихідна інформація для надання зовнішнім користувачам

За результатами застосування статистико-математичних моделей зовнішнім користувачам інформація надається фактично в тому ж обсязі, що і за результатами прямого оцінювання. При цьому замість прямих оцінок визначених показників бідності за регіонами

надаються уточнені (композиційні) оцінки, замість коефіцієнтів варіації – відповідні відносні похибки (величини відношення кореня квадратного від середньоквадратичної похибки до композиційної оцінки у відсотках). Розрахунки обов'язково супроводжуються методологічними поясненнями щодо оцінювання показників бідності з використанням методів статистико-математичного моделювання.

При підготовці інформації слід враховувати та звертати увагу користувачів, що композиційне оцінювання показників бідності за регіонами вимагає перерахунку показників рівня України та, наприклад, рівня економічних районів (див. формули 9–11). Для економічних районів уточнені оцінки показників бідності розраховуються за аналогічними формулами, але з урахування інформації лише по тих регіонах, які входять до складу відповідних економічних районів.

Результати оцінювання показників бідності на основі статистико-математичних моделей доцільно представляти у вигляді таблиць, діаграм та графіків. Нижче наведені рекомендовані макети таблиць (див. табл. 3–5) для представлення інформації щодо показника рівня бідності, визначеного за критерієм національної межі.

Таблиця 3

Уточнені оцінки рівня бідності, визначених за національною межею, та характеристики їх надійності за регіонами України у 20__ році

	Уточнена оцінка рівня бідності за попередній рік (довідково), %	Уточнена оцінка рівня бідності, %	Відносна похибка оцінювання, <i>RRMSE</i> , %	Довірчий інтервал, %
Україна				
АР Крим				
Вінницька область				
Волинська область				
Дніпропетровська область				
Донецька область				
Житомирська область				
Закарпатська область				
Запорізька область				
Івано-Франківська область				
Київська область				
Кіровоградська область				
Луганська область				
Львівська область				
Миколаївська область				
Одеська область				
Полтавська область				
Рівненська область				

Сумська область				
Тернопільська область				
Харківська область				
Херсонська область				
Хмельницька область				
Черкаська область				
Чернівецька область				
Чернігівська область				
м. Київ				
м. Севастополь				

Таблиця 4

Уточнені оцінки рівня бідності за національною межею та характеристики їх надійності за економічними районами у 20__ році

	Уточнена оцінка рівня бідності за попередній рік (довідково), %	Уточнена оцінка рівня бідності, %	Відносна похибка оцінювання, <i>RRMSE</i> , %	Довірчий інтервал, %
Східний				
Донецький				
Придніпровський				
Причорноморський				
Подільський				
Центральний				
Карпатський				
Поліський				

Для візуалізації результатів непрямого оцінювання показників бідності доцільно представити користувачам карти бідності для регіонів та економічних районів України (рис. 1, 2).

Аналогічні таблиці та карти наводяться по кожному показнику бідності, для якого визначаються уточнені оцінки, зокрема для рівня бідності, визначеного за регіональною межею, та для рівня бідності, визначеного за розміром прожиткового мінімуму.



Рис.1. Карта бідності за показником рівня бідності за регіональною межею за регіонами України у 20__ році.

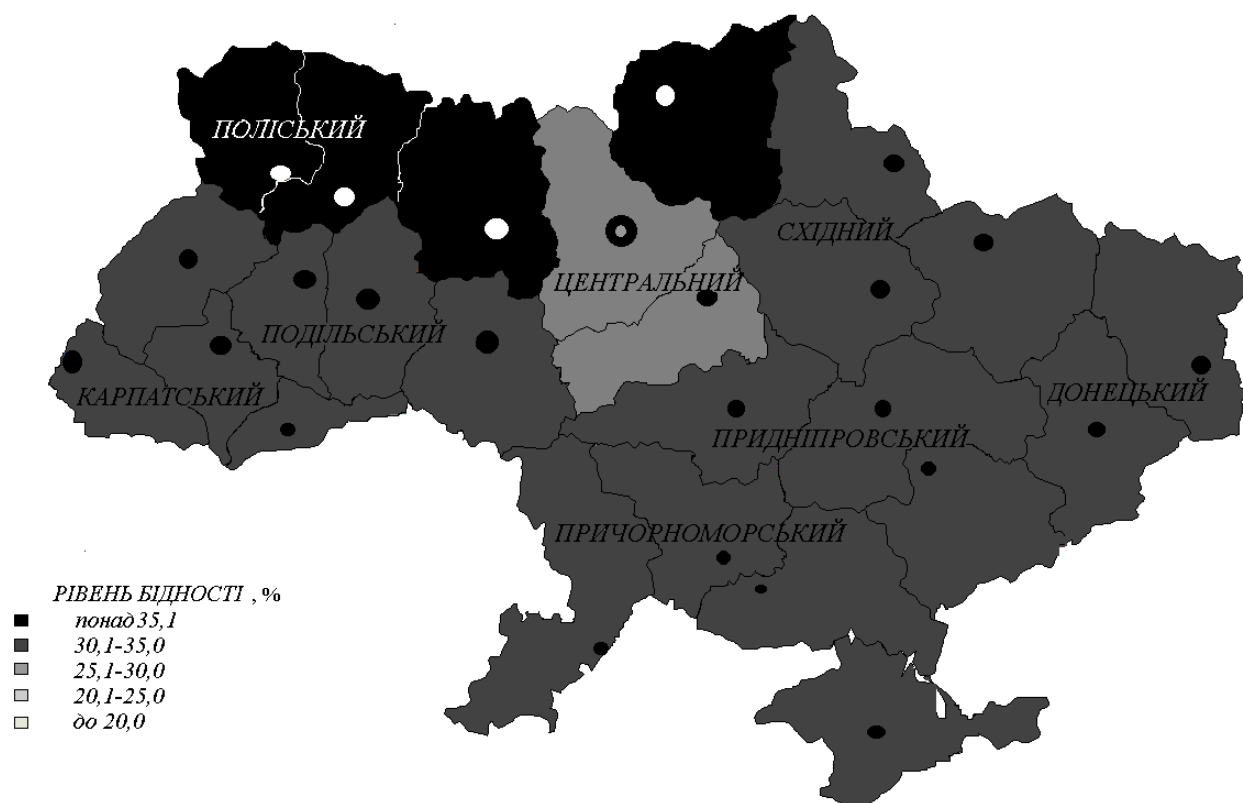


Рис.2. Карта бідності за показником рівня бідності за національною межею за економічними районами України у 20__ році.

Для зовнішніх користувачів також доцільно навести по кожному показнику бідності основні характеристики моделі (див. табл. 5).

Таблиця 5

**Основні характеристики статистико-математичної моделі
для оцінювання рівня бідності _____**

Назва показника	Оцінка показника
Коефіцієнт регресії (вільний член), $\hat{\beta}_0$	
Коефіцієнт регресії, $\hat{\beta}_1$	
Коефіцієнт регресії*, $\hat{\beta}_2$	
Коефіцієнт детермінації, R^2	
F -відношення	
Дисперсії синтетичної оціночної функції, $\hat{\sigma}_v^2$	

* Для визначеності прийнято, що синтетична оціночна функція побудована на основі двох зовнішніх змінних:

$$\hat{\theta}_k^{(S)} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 .$$

Характеристики, наведені у табл. 5, які мають супроводжуватися відповідними методологічними коментарями, дозволять користувачу мати уявлення щодо рівня адекватності синтетичної моделі та краще зрозуміти, яким чином було уточнено показник бідності на основі композиційної оціночної функції.

Додатки

Додаток 1

Надійність прямих оцінок визначених показників бідності для національного та регіонального рівня (ОУЖД, 2007)

%

Регіон	Рівень бідності за національною межею, $\hat{P}_{1k}$	Коефіцієнт варіації, $CV(\hat{P}_{1k})$	Рівень бідності за регіональною межею, $\hat{P}_{2k}$	Коефіцієнт варіації, $CV(\hat{P}_{2k})$	Рівень бідності за прожитковим мінімумом, $\hat{P}_{3k}$	Коефіцієнт варіації, $CV(\hat{P}_{3k})$
Україна	27,27	2,47	32,30	2,04	12,86	4,72
АР Крим	29,20	13,04	37,33	10,82	20,58	23,05
Вінницька	23,03	14,43	24,63	14,88	17,29	18,00
Волинська	39,87	13,18	38,31	13,75	25,66	16,23
Дніпропетровська	27,85	11,26	32,94	9,43	12,35	15,32
Донецька	22,43	10,07	30,61	8,37	6,97	15,76
Житомирська	35,61	9,26	40,12	8,68	16,01	18,42
Закарпатська	24,28	9,26	27,83	11,79	10,47	32,51
Запорізька	21,84	12,91	24,32	12,36	12,28	17,15
Івано-Франківська	16,88	22,48	21,64	21,42	9,25	28,13
Київська	29,98	14,12	42,24	9,72	9,47	20,04
Кіровоградська	45,57	10,22	46,96	11,42	25,53	21,09
Луганська	39,34	7,24	42,59	6,90	12,56	13,00
Львівська	29,79	8,82	34,02	8,44	13,73	15,21
Миколаївська	19,63	14,93	20,42	14,07	14,10	18,84
Одеська	34,61	11,83	42,58	9,55	13,68	15,43
Полтавська	27,47	12,41	31,84	11,72	11,03	19,11
Рівненська	34,78	10,81	35,76	10,92	14,24	24,28
Сумська	30,87	11,84	34,60	10,52	12,94	20,71
Тернопільська	33,30	10,74	36,97	6,94	19,89	21,90
Харківська	28,18	12,00	32,77	10,49	8,74	24,55
Херсонська	27,44	10,52	26,42	10,24	17,30	26,60
Хмельницька	34,69	7,39	35,07	7,35	15,24	11,76
Черкаська	16,29	12,98	24,28	10,69	7,44	22,03
Чернівецька	28,10	11,66	38,96	9,68	19,62	20,78
Чернігівська	27,26	15,22	30,33	11,62	17,21	21,12
м. Київ	8,00	25,36	16,75	17,90	5,46	39,83
м. Севастополь	7,91	47,17	18,32	33,21	0,28	115,75

**Надійність прямих оцінок визначених показників бідності
для національного та регіонального рівня (ОУЖД, 2008)**

%

Регіон	Рівень бідності за національ- ною межею, $\hat{P}_{1k}$	Коефіцієнт варіації, $CV(\hat{P}_{1k})$	Рівень бідності за регіо- нальною межею, $\hat{P}_{2k}$	Коефіцієнт варіації, $CV(\hat{P}_{2k})$	Рівень бідності за прожитко- вим мінімумом, $\hat{P}_{3k}$	Коефіцієнт варіації, $CV(\hat{P}_{3k})$
Україна	26,97	2,41	32,62	2,15	7,10	5,40
АР Крим	24,89	9,21	34,36	6,96	7,80	26,30
Вінницька	24,30	13,86	27,14	12,71	6,90	21,70
Волинська	47,30	10,21	47,74	9,93	15,60	22,40
Дніпропетровська	27,22	13,01	32,25	11,48	5,60	22,60
Донецька	25,58	9,66	32,16	7,95	4,10	24,40
Житомирська	34,22	14,72	39,14	11,29	9,60	22,30
Закарпатська	20,30	22,07	24,00	19,93	8,30	39,00
Запорізька	24,39	15,47	30,35	12,91	6,80	33,70
Івано-Франківська	20,09	18,93	24,44	14,81	3,50	43,10
Київська	21,17	23,63	35,13	16,69	5,00	51,90
Кіровоградська	47,09	9,14	47,09	9,14	21,30	14,90
Луганська	29,20	8,74	34,32	9,77	6,40	21,80
Львівська	29,25	8,49	38,66	8,12	7,10	19,10
Миколаївська	19,81	16,35	19,81	16,35	6,70	18,60
Одеська	38,48	12,04	43,99	10,24	10,70	21,90
Полтавська	27,47	12,36	32,48	10,60	6,90	28,40
Рівненська	39,28	11,51	39,61	11,93	12,90	19,00
Сумська	29,07	12,16	30,25	11,85	4,90	25,10
Тернопільська	42,77	10,93	46,50	10,10	10,10	39,30
Харківська	20,59	9,90	31,23	9,00	2,80	28,20
Херсонська	32,96	12,80	32,96	12,80	12,30	18,90
Хмельницька	29,79	7,79	32,79	8,71	11,30	22,40
Черкаська	18,37	20,64	21,44	22,08	6,20	24,90
Чернівецька	29,60	15,88	35,78	12,91	7,00	32,10
Чернігівська	28,62	11,88	31,13	10,88	10,00	17,50
м. Київ	9,63	17,11	19,36	11,09	1,10	72,90
м. Севастополь	2,70	73,05	14,59	56,86	0,80	62,30

Опис узагальненого методу найменших квадратів

Коефіцієнти регресії при побудові синтетичної оціночної функції визначаються за формулою [4]:

$$\hat{\beta} = \left[\sum_{k=1}^K \frac{x_k x_k^T}{\psi_k + \sigma_v^2} \right]^{-1} \left[\sum_{k=1}^K \frac{x_k \hat{\theta}_k}{\psi_k + \sigma_v^2} \right],$$

де $\hat{\beta}$ – вектор оцінок коефіцієнтів регресії; символ $\square^{-1}$ позначає обернену матрицю.

Враховуючи, що дійсне значення похибки моделі σ_v^2 є невідомим, замість нього використовують наближене значення $\hat{\sigma}_v^2$ (див. додаток 3). Відповідно, при цьому коефіцієнти регресії є також оцінками дійсних коефіцієнтів β .

Опис методу моментів

Величина $\hat{\sigma}_v^2$ може визначатися за методом моментів із ітераційної процедури, запропонованої Феєм та Херріотом [4, 6]:

$$\begin{aligned} \hat{\sigma}_v^{2(a+1)} = \hat{\sigma}_v^{2(a)} - \frac{1}{\sum_{k=1}^K (\hat{\theta}_k - x_k^T \hat{\beta}(\hat{\sigma}_v^{2(a)}))^2 / (\psi_k + \hat{\sigma}_v^{2(a)})^2} \times \\ \times (K - p + \sum_{k=1}^K (\hat{\theta}_k - x_k^T \hat{\beta}(\hat{\sigma}_v^{2(a)}))^2 / (\psi_k + \hat{\sigma}_v^{2(a)})^2) \end{aligned},$$

де p – кількість корелят у моделі.

При здійсненні розрахунків приймається, що $\hat{\sigma}_v^{2(0)} = 0$. Кількість необхідних ітерацій звичайно не перевищує 10.

Приклад побудови прототипів моделей

У додатку наведено приклад побудови прототипу моделі для підвищення надійності оцінювання рівня бідності, визначеного за національною межею, для регіонів України.

1. Дослідження джерел даних

У відповідності з рекомендаціями розділу 3 методологічних положень насамперед доцільно приділити увагу результатам ОУЖД. Кореляційні матриці для розподілів рівнів бідності за національною межею за регіонами та розподілів логарифмів від них за 2006–2008 роки наведені у табл. 1–3. Вихідні дані, на основі яких розраховувались коефіцієнти кореляції, представлені у додатку 5. Розподіли логарифмів рівнів бідності аналізуються для регіонів як із м. Севастополь, так і без нього, оскільки оцінки показників для м. Севастополь, які є дуже ненадійними, можуть помітно впливати на величини коефіцієнтів кореляції та, відповідно, на параметри композиційної оціночної функції. За наведеними даними, між суміжними роками кореляція розподілів рівнів бідності, що розглядаються, за регіонами є достатньо високою (коефіцієнт кореляції перевищує 0,85). З часом щільність взаємозв'язку дещо знижується, але несуттєво: між показниками за 2006 та 2008 роки коефіцієнт кореляції перевищує 0,83, що є ознакою досить щільного лінійного взаємозв'язку. Для логарифмів рівнів бідності значення коефіцієнтів кореляції є навіть більшими, ніж для самих рівнів. При цьому слід зауважити, що кореляційна матриця для логарифмів рівнів бідності, розрахована без урахування м. Севастополь (див. табл. 3), свідчить про зниження рівня взаємозв'язку між розподілами з часом, що є більш вірогідним, ніж його зростання, як це демонструє кореляційна матриця для логарифмів рівнів бідності, розрахована з урахуванням м. Севастополь (див. табл. 2).

Ураховуючи щорічну повну ротацію домогосподарств у вибірці ОУЖД, слід зробити висновок, що такий взаємозв'язок потенційно може бути використаний для ефективного уточнення оцінок рівня бідності за національною межею за регіонами України.

Таблиця 1. Коефіцієнти кореляції Пірсона між розподілами рівня бідності за національною межею за регіонами (ОУЖД, 2006–2008 роки)*

	Рівень бідності, 2006	Рівень бідності, 2007	Рівень бідності, 2008
Рівень бідності, 2006	1	0,859	0,832
Рівень бідності, 2007		1	0,884
Рівень бідності, 2008			1

*Тут і далі досліджуються взаємозв'язки не рівнів показників, а їх розподілів за регіонами.

Таблиця 2. Коефіцієнти кореляції Пірсона між розподілами логарифмів (натуральних) рівня бідності за національною межею за регіонами (ОУЖД, 2006–2008 роки)

	Логарифм (натуральний) рівня бідності, 2006	Логарифм (натуральний) рівня бідності, 2007	Логарифм (натуральний) рівня бідності, 2008
Логарифм (натуральний) рівня бідності, 2006	1	0,861	0,936
Логарифм (натуральний) рівня бідності, 2007		1	0,892
Логарифм (натуральний) рівня бідності, 2008			1

Таблиця 3. Коефіцієнти кореляції Пірсона між розподілами логарифмів (натуральних) рівня бідності за національною межею за регіонами без м. Севастополь (ОУЖД, 2006–2008 роки)

	Логарифм (натуральний) рівня бідності, 2006	Логарифм (натуральний) рівня бідності, 2007	Логарифм (натуральний) рівня бідності, 2008
Логарифм (натуральний) рівня бідності, 2006	1	0,870	0,810
Логарифм (натуральний) рівня бідності, 2007		1	0,890
Логарифм (натуральний) рівня бідності, 2008			1

Дані, що ілюструють надійність оцінювання рівня бідності за національною межею за економічними районами, наведені у табл. 4. Як видно з наведених даних, надійність оцінок цього показника майже по всіх економічних районах практично відповідає критерію надійності, встановленому для регіонів, а саме – величина коефіцієнта варіації вибірових оцінок CV не перевищує 10% (виключенням є Центральний економічний район, для якого у

2008 році $CV = 12,69$). Аналогічна ситуація спостерігається і для оцінок рівня бідності за регіональною межею. Щодо оцінювання частки населення з еквівалентними доходами, нижчими прожиткового мінімуму, то надійність цього показника не для всіх економічних районів задовольняє критерію надійності і, таким чином, можливість використання інформації цього рівня є достатньо обмеженою.

За результатами виконаних досліджень встановлено, що використання оцінок показників для вищих рівнів агрегації даних (національний рівень, економічні райони) при побудові синтетичних оцінок не є статистично ефективним.

Таблиця 4. Надійність оцінювання рівня бідності за національною межею за економічними районами (ОУЖД, 2007–2008)

	Рівень бідності, %, 2007	Коефіцієнт варіації, % 2007	Рівень бідності, %, 2008	Коефіцієнт варіації, % 2008
Україна	27,3	2,5	27,0	2,41
Східний	28,6	7,39	24,4	6,53
Донецький	28,3	6,31	26,8	6,81
Придніпровський	29,1	7,23	29,8	7,83
Причорноморський	28,0	6,75	28,8	6,83
Подільський	29,6	6,55	31,0	6,29
Центральний	16,6	10,39	15,1	12,69
Карпатський	25,5	5,95	25,4	6,98
Поліський	34,3	6,01	37,0	6,09

Для розподілу оцінок рівня бідності за національною межею за регіонами встановлено кореляційний взаємозв'язок із розподілом кінцевих споживчих витрат домогосподарств у розрахунку на одну особу за даними національних рахунків (див. табл. 5). Вихідні дані, на основі яких розраховувались коефіцієнти кореляції, представлені у додатках 5, 6.

Таким чином, для підвищення надійності оцінювання рівня бідності за національною межею, корисним може бути використання розподілу за регіонами величини кінцевих споживчих витрат домогосподарств у розрахунку на одну особу. Особливу увагу слід приділити синтетичним оцінкам, що побудовані з використанням натуральних логарифмів від показників. При побудові реальних моделей на основі запропонованих прототипів та їх апробації необхідно дослідити доцільність урахування м. Севастополь в якості окремого регіону.

Таблиця 5. Коефіцієнти кореляції Пірсона між розподілами рівнів бідності за національною межею і логарифмів (натуральних) від них та кінцевими споживчими витратами домогосподарств у розрахунку на одну особу та логарифмів (натуральних) від них за регіонами України

	Рівень бідності за національною межею (2007), %	Логарифм (натуральний) від рівня бідності за національною межею (2007), %	Рівень бідності за національною межею (2008), %	Логарифм (натуральний) від рівня бідності за національною межею (2008), %
Кінцеві споживчі витрати домогосподарств у розрахунку на одну особу (2007), грн.	-0,537	-0,656	-0,488	-0,507
Логарифм (натуральний) від кінцевих споживчих витрат домогосподарств у розрахунку на одну особу (2007)	-0,563	-0,667 (-0,666)*	-0,545	-0,570
Кінцеві споживчі витрати домогосподарств у розрахунку на одну особу (2008), грн.	-0,533	-0,653	-0,494	-0,511
Логарифм (натуральний) від кінцевих споживчих витрат домогосподарств у розрахунку на одну особу (2008)	-0,557	-0,665	-0,547	-0,572 (-0,668)*

* Коефіцієнти кореляції розподілів логарифмів відповідних показників без урахування м. Севастополь

2. Побудова моделі

При ілюстрації особливостей побудови моделей, в якості приклада розглядаються результати уточнення оцінок рівнів бідності, визначених за національною межею, для регіонів України за даними ОУЖД 2008 та 2007 років. При цьому в якості зовнішньої інформації використовуються такі показники:

- прямі оцінки рівнів бідності за національною межею для регіонів за результатами ОУЖД 2006 року при композиційному оцінюванні рівня бідності у 2007 році і прямі оцінки рівнів бідності за національною межею для регіонів за результатами ОУЖД 2007 року при композиційному оцінюванні рівня бідності у 2008 році;
- розподіл за регіонами обсягу кінцевих споживчих витрат домогосподарств у розрахунку на одну особу за даними національних рахунків за 2007 рік при оцінюванні рівня бідності у

2007 році, і розподіл за регіонами обсягу кінцевих споживчих витрат домогосподарств у розрахунку на одну особу за даними національних рахунків за 2008 рік при оцінюванні рівня бідності у 2008 році.

Для спрощення ілюстрації тут не наведено прикладів використання різних за складом та кількістю показників сукупностей зовнішніх змінних, хоча при апробації запропонованого прототипу моделі розрахунки виконувались для різних варіантів синтетичних оціночних функцій, представлених одно-, двох- і трьохфакторними рівняннями регресії, коли в ролі факторних змінних виступали різні показники, розглянуті у [10]. Представлені результати відповідають варіанту з найменшою величиною середньоквадратичних похибок композиційних оцінок MSE рівнів бідності за національною межею за регіонами.

Дані, наведені у табл. 6 і 7, ілюструють оцінки за регіонами рівнів бідності, визначених за національною межею, та відносні похибки оцінювання.

При композиційному оцінюванні регіональних рівнів бідності за національною межею у 2007 році синтетична оціночна функція визначалася за формулою:

$$\hat{\theta}_{k,2007}^{(S)} = \hat{P}_{k,2007}^{(S)} = 12.085 + 0.649 \cdot \hat{P}_{k,2006} - 0.385 \cdot 10^{-3} \cdot x_{2k}, \quad (1)$$

де $\hat{\theta}_{k,2007}^{(S)}$ – узагальнений вигляд синтетичної оцінки показника бідності у 2007 році для k -го регіону; $\hat{P}_{k,2007}^{(S)}$ – синтетична оцінка рівня бідності за національною межею у 2007 році для k -го регіону; $\hat{P}_{k,2006}$ – пряма оцінка рівня бідності за національною межею у 2006 році для k -го регіону; x_{2k} – обсяг кінцевих споживчих витрат домогосподарств у розрахунку на одну особу за даними національних рахунків у 2007 році для k -го регіону.

При композиційному оцінюванні регіональних рівнів бідності за національною межею у 2008 році синтетична оціночна функція визначалася за формулою:

$$\hat{\theta}_{k,2008}^{(S)} = \hat{P}_{k,2008}^{(S)} = 1.089 + 0.949 \cdot \hat{P}_{k,2007} - 0.024 \cdot 10^{-3} \cdot x_{2k}, \quad (2)$$

де $\hat{\theta}_{k,2008}^{(S)}$ – узагальнений вигляд синтетичної оцінки показника бідності у 2008 році для k -го регіону; $\hat{P}_{k,2008}^{(S)}$ – синтетична оцінка рівня бідності за національною межею у 2008 році для k -го регіону; $\hat{P}_{k,2007}$ – пряма оцінка рівня бідності за національною межею у 2007 році для k -го регіону; x_{2k} – обсяг кінцевих споживчих витрат домогосподарств у розрахунку на одну особу за даними національних рахунків у 2008 році для k -го регіону.

Композиційна оціночна функція для k -го регіону визначається за формулою (6).

Таблиця 6. Результати композиційного оцінювання рівнів бідності за національною межею у 2007 році за регіонами України на основі прямих оцінок рівнів бідності за національною межею 2007 року, прямих оцінок рівнів бідності за національною межею 2006 року, розподілу за регіонами обсягу кінцевих споживчих витрат домогосподарств у розрахунку на одну особу за даними національних рахунків за 2007 рік

Регіон	Пряма оцінка показника, $\hat{P}_k$, %	Коефіцієнт варіації прямої оцінки, CV_k , %	Синтетична оцінка, $\hat{P}_k^{(S)}$, %	Ваговий коефіцієнт, $\hat{\gamma}_k$	Композиційна оцінка, $\hat{P}_k^{(H)}$, %	Відносна похибка, $RRMSE(\hat{P}_k^{(H)})$, %
АР Крим	29,20	13,01	28,74	0,439	28,94	9,45
Вінницька область	23,03	14,33	23,11	0,509	23,07	11,21
Волинська область	39,87	13,29	42,13	0,287	41,48	8,39
Дніпропетровська область	27,85	11,13	26,57	0,540	27,26	9,03
Донецька область	22,43	10,25	24,91	0,681	23,22	8,74
Житомирська область	35,61	9,27	33,33	0,509	34,49	7,48
Закарпатська область	24,28	9,06	24,01	0,700	24,20	8,19
Запорізька область	21,84	12,82	25,91	0,590	23,51	9,86
Івано-Франківська область	16,88	22,51	17,84	0,439	17,42	16,84
Київська область	29,98	14,01	34,72	0,390	32,87	8,97
Кіровоградська область	45,57	10,31	36,26	0,338	39,41	7,82
Луганська область	39,34	7,12	33,71	0,590	37,03	6,33
Львівська область	29,79	8,73	25,15	0,625	28,05	7,90
Миколаївська область	19,63	14,77	22,44	0,573	20,83	11,50
Одеська область	34,61	11,85	30,85	0,402	32,36	8,89
Полтавська область	27,47	12,38	24,03	0,494	25,73	10,13
Рівненська область	34,78	10,93	32,64	0,439	33,58	8,22
Сумська область	30,87	11,99	29,17	0,452	29,94	9,04
Тернопільська область	33,30	10,81	32,12	0,466	32,67	8,23
Харківська область	28,18	12,07	25,04	0,494	26,59	9,75
Херсонська область	27,44	10,57	37,28	0,573	31,64	7,76
Хмельницька область	34,69	7,49	30,33	0,626	33,06	6,70
Черкаська область	16,29	12,89	27,25	0,719	19,37	9,78
Чернівецька область	28,10	11,74	27,65	0,509	27,88	9,17
Чернігівська область	27,26	15,04	26,16	0,402	26,60	10,68
м. Київ	8,00	25,00	7,39	0,738	7,84	26,24
м. Севастополь	7,91	46,78	8,66	0,452	8,32	36,45

Таблиця 7. Результати композиційного оцінювання рівнів бідності за національною межею у 2008 році за регіонами України на основі прямих оцінок рівнів бідності за національною межею 2008 року, прямих оцінок рівнів бідності за національною межею 2007 року, розподілу за регіонами обсягу кінцевих споживчих витрат домогосподарств у розрахунку на одну особу за даними національних рахунків за 2008 рік

	Пряма оцінка показника, $\hat{P}_k$, %	Коефіцієнт варіації прямої оцінки, CV_k , %	Синтетична оцінка, $\hat{P}_k^{(S)}$, %	Ваговий коефіцієнт, $\hat{\gamma}_k$	Композиційна оцінка, $\hat{P}_k^{(H)}$, %	Відносна похибка, $RRMSE(\hat{P}_k^{(H)})$, %
АР Крим	24,89	9,21	28,41	0,665	26,07	7,74
Вінницька область	24,30	13,86	22,61	0,480	23,42	11,00
Волинська область	47,30	10,21	38,60	0,310	41,30	7,50
Дніпропетровська область	27,22	13,01	27,15	0,455	27,18	9,61
Донецька область	25,58	9,66	22,03	0,631	24,27	8,80
Житомирська область	34,22	14,72	34,57	0,292	34,47	8,80
Закарпатська область	20,30	22,07	23,85	0,343	22,63	12,88
Запорізька область	24,39	15,47	21,44	0,423	22,69	11,94
Івано-Франківська область	20,09	18,93	16,82	0,419	18,19	15,68
Київська область	21,17	23,63	29,14	0,295	26,79	11,12
Кіровоградська область	47,09	9,14	44,04	0,361	45,14	6,93
Луганська область	29,20	8,74	38,13	0,616	32,63	6,83
Львівська область	29,25	8,49	29,03	0,628	29,17	7,32
Миколаївська область	19,81	16,35	19,37	0,499	19,59	13,02
Одеська область	38,48	12,04	33,53	0,327	35,15	8,50
Полтавська область	27,47	12,36	26,82	0,476	27,13	9,46
Рівненська область	39,28	11,51	33,81	0,338	35,66	8,20
Сумська область	29,07	12,16	30,10	0,456	29,63	8,86
Тернопільська область	42,77	10,93	32,41	0,324	35,77	8,23
Харківська область	20,59	9,90	27,42	0,716	22,53	8,22
Херсонська область	32,96	12,80	26,82	0,370	29,09	9,68
Хмельницька область	29,79	7,79	33,70	0,660	31,12	6,60
Черкаська область	18,37	20,64	16,21	0,421	17,12	16,49
Чернівецька область	29,60	15,88	27,45	0,321	28,14	10,38
Чернігівська область	28,62	11,88	26,64	0,475	27,58	9,32
м. Київ	9,63	17,11	7,57	0,796	9,21	18,49
м. Севастополь	2,70	73,05	8,10	0,728	4,17	45,05

Характеристики адекватності синтетичних моделей:

2007 рік

$$R^2 = 0,763; F = 38,587 (F_{кр} = 4,58).$$

2008 рік

$$R^2 = 0,782; F = 43,139 (F_{кр} = 4,58).$$

Для оцінювання рівня бідності за національною межею на національному рівні застосовується формула, аналогічна (9):

$$\hat{P}_U^{(H)} = \frac{\sum_{k=1}^K N_k \hat{P}_k^{(H)}}{\sum_{k=1}^K N_k}, \quad (3)$$

де $\hat{P}_U^{(H)}$ – уточнена оцінка рівня бідності за національною межею на національному рівні;
 $\hat{P}_k^{(H)}$ – композиційна оцінка рівня бідності за національною межею для k -го регіону; N_k – середньорічна чисельність населення k -го регіону.

Середньоквадратичні похибки оцінок показників бідності на національному рівні $MSE(\hat{P}_U^{(H)})$ визначаються на основі середньоквадратичних похибок оцінок показників для регіонів за формулою, аналогічною (10):

$$MSE(\hat{P}_U^{(H)}) = \frac{1}{N^2} \sum_{k=1}^K N_k^2 MSE(\hat{P}_k^{(H)}). \quad (4)$$

На основі величини $MSE(\hat{P}_U^{(H)})$ розраховується значення $RRMSE(\hat{P}_U^{(H)})$: відношення кореня квадратного від середньоквадратичної похибки до композиційної оцінки показника бідності за формулою, аналогічною (11).

Розрахована на основі формули (3) для 2007 року уточнена оцінка рівня бідності за національною межею на національному рівні складає $\hat{P}_U^{(H)} = 26,9\%$. При цьому $RRMSE(\hat{P}_U^{(H)}) = 1,95\%$. Пряма оцінка цього показника: $\hat{P}_U = 27,3\%$ при $CV(\hat{P}_U) = 2,51\%$.

Для 2008 року уточнена оцінка рівня бідності за національною межею на національному рівні складає $\hat{P}_U^{(H)} = 26,5\%$. При цьому $RRMSE(\hat{P}_U^{(H)}) = 1,94\%$. Пряма оцінка цього показника: $\hat{P}_U = 27,0\%$ при $CV(\hat{P}_U) = 2,41\%$.

**Рівень бідності за національною межею та логарифми від них за регіонами України
(ОУЖД 2006–2008 роки)**

Регіон	Рівень бідності (2006), %	Логарифм (натуральний) рівня бідності (2006)	Рівень бідності, (2007), %	Логарифм (натуральний) рівня бідності (2007)	Рівень бідності, (2008), %	Логарифм (натуральний) рівня бідності (2008)
АР Крим	30,83	3,4285	29,20	3,3742	24,89	3,2145
Вінницька	21,29	3,0582	23,03	3,1368	24,30	3,1905
Волинська	50,67	3,9253	39,87	3,6856	47,30	3,8565
Дніпропетровська	27,63	3,3189	27,85	3,3268	27,22	3,3040
Донецька	24,84	3,2125	22,43	3,1104	25,58	3,2418
Житомирська	36,68	3,6022	35,61	3,5726	34,22	3,5328
Закарпатська	22,27	3,1032	24,28	3,1897	20,30	3,0106
Запорізька	26,60	3,2809	21,84	3,0837	24,39	3,1942
Івано-Франківська	12,71	2,5424	16,88	2,8261	20,09	3,0002
Київська	39,90	3,6864	29,98	3,4005	21,17	3,0526
Кіровоградська	41,39	3,7230	45,57	3,8192	47,09	3,8521
Луганська	37,10	3,6136	39,34	3,6722	29,20	3,3742
Львівська	24,78	3,2100	29,79	3,3942	29,25	3,3759
Миколаївська	20,64	3,0272	19,63	2,9771	19,81	2,9862
Одеська	34,89	3,5522	34,61	3,5441	38,48	3,6501
Полтавська	23,06	3,1381	27,47	3,3131	27,47	3,3131
Рівненська	35,46	3,5684	34,78	3,5490	39,28	3,6707
Сумська	30,14	3,4059	30,87	3,4298	29,07	3,3697
Тернопільська	34,26	3,5340	33,30	3,5056	42,77	3,7558
Харківська	26,01	3,2585	28,18	3,3386	20,59	3,0248
Херсонська	43,22	3,7663	27,44	3,3120	32,96	3,4953
Хмельницька	32,43	3,4791	34,69	3,5465	29,79	3,3942
Черкаська	27,91	3,3290	16,29	2,7906	18,37	2,9107
Чернівецька	28,09	3,3354	28,10	3,3358	29,60	3,3878
Чернігівська	25,65	3,2445	27,26	3,3054	28,62	3,3541
м. Київ	8,79	2,1736	8,00	2,0794	9,63	2,2649
м. Севастополь	1,42	0,3507	7,91	2,0681	2,70	0,9933

Кінцеві споживчі витрати домогосподарств у розрахунку на одну особу та логарифми від них за регіонами України (2007, 2008 роки)

Регіон	Кінцеві споживчі витрати домогосподарств у розрахунку на одну особу (2007), грн.	Логарифм (натуральний) кінцевих споживчих витрат домогосподарств у розрахунку на одну особу (2007)	Кінцеві споживчі витрати домогосподарств у розрахунку на одну особу (2008), грн.	Логарифм (натуральний) кінцевих споживчих витрат домогосподарств у розрахунку на одну особу (2008)
АР Крим	8723,0	9,0737	15900,4	9,6741
Вінницька	7277,0	8,8925	14011,8	9,5477
Волинська	7391,0	8,9080	13842,5	9,5355
Дніпропетровська	8996,0	9,1045	16147,7	9,6895
Донецька	8618,0	9,0616	15269,4	9,6336
Житомирська	6663,0	8,8043	13192,8	9,4874
Закарпатська	6612,0	8,7966	12256,6	9,4138
Запорізька	8961,0	9,1006	15666,9	9,6593
Івано-Франківська	6480,0	8,7765	12459,8	9,4303
Київська	8480,0	9,0455	16911,0	9,7357
Кіровоградська	6979,0	8,8507	13147,9	9,4840
Луганська	6437,0	8,7698	12236,3	9,4122
Львівська	7862,0	8,9698	14085,0	9,5529
Миколаївська	7900,0	8,9746	14787,0	9,6015
Одеська	10078,0	9,2181	17376,5	9,7629
Полтавська	7849,0	8,9681	14165,1	9,5585
Рівненська	6405,0	8,7648	12437,7	9,4285
Сумська	6462,0	8,7737	12505,0	9,4339
Тернопільська	5740,0	8,6552	11381,4	9,3397
Харківська	10240,0	9,2341	18266,2	9,8128
Херсонська	7421,0	8,9121	13406,1	9,5035
Хмельницька	7268,0	8,8912	13456,7	9,5072
Черкаська	7685,0	8,9470	14146,1	9,5572
Чернівецька	6951,0	8,8466	12889,5	9,4642
Чернігівська	6682,0	8,8072	13336,9	9,4983
м. Київ	27027,0	10,2046	44054,6	10,6932
м. Севастополь	11321,0	9,3344	19828,9	9,8949

Проект методологічних пояснень щодо оцінювання показників бідності з використанням методів математико-статистичного моделювання

У статистичному збірнику представлені показники бідності для національного та регіонального рівнів, отримані на основі застосування спеціального підходу до оцінювання показників на основі методів статистичного моделювання. Метою застосування цих методів є підвищення надійності оцінювання показників бідності на регіональному рівні.

Загальний опис проблеми

Показники бідності в Україні вимірюються на основі даних щоквартального державного вибіркового обстеження умов життя домогосподарств (далі – ОУЖД) у відповідності до діючих методичних та нормативних документів. Це обстеження проводиться щороку на постійній основі, починаючи з 1999 р., і відповідає міжнародним стандартам.

Статистичний аналіз характеристик надійності показників бідності, визначених за даними ОУЖД, засвідчив, що прямі оцінки показників на рівні окремих регіонів вимірюються з недостатньою точністю. Це призводить до нестабільності оцінок у динаміці, ускладнює порівняння показників та в цілому обмежує можливість ефективного моніторингу бідності в Україні.

Ефективним способом підвищення надійності показників бідності для різних рівнів агрегації даних є застосування спеціальних методологічних підходів до їх оцінювання, що базуються на методах непрямого оцінювання. Застосування таких методів, як правило, здійснюється для оцінювання лише найбільш важливих показників, оскільки вимагає реалізації додаткових складних технологічних процедур [1, 2].

Система показників

У публікації представлені результати непрямого оцінювання таких показників бідності:

1) Рівень бідності за національною межею $P_k^{(1)}$ – питома вага сімей (домогосподарств), у яких рівень споживання (доходів) на одну особу є нижчим від визначеної межі бідності. Рівень бідності розраховується за формулою:

$$P_k^{(1)} = \frac{Q_k^{(1)}}{N_k} \cdot 100\% , \quad (1)$$

де $Q_k^{(1)}$ – чисельність населення k -го регіону, що визнається бідним відповідно до національної межі бідності; N_k – загальна чисельність населення k -го регіону.

Межа бідності визначається на підставі відносного критерію зарахування різних верств населення до категорії бідних, який розраховується за фіксованою часткою середньодушового доходу (витрат) – 75% медіанного рівня сукупних доходів (витрат) у розрахунку на умовного дорослого.

2) Рівень бідності за регіональною межею $P_k^{(2)}$ – частка населення регіону, доходи яких є нижчими за розраховану регіональну межу бідності. Цей рівень бідності розраховується за формулою:

$$P_k^{(2)} = \frac{Q_k^{(2)}}{N_k} \cdot 100\% , \quad (2)$$

де $Q_k^{(2)}$ – чисельність населення k -го регіону, що визнається бідним відповідно до регіональної межі бідності; N_k – загальна чисельність населення k -го регіону.

3) Рівень бідності за розміром прожиткового мінімуму $P_k^{(3)}$ – частка населення з загальними середньомісячними еквівалентними доходами, нижчими за прожитковий мінімум у середньому на одну особу. Цей рівень бідності визначається за формулою:

$$P_k^{(3)} = \frac{Q_k^{(3)}}{N_k} \cdot 100\% , \quad (3)$$

де $Q_k^{(3)}$ – чисельність населення k -го регіону з загальними середньомісячними еквівалентними доходами, нижчими за прожитковий мінімум; N_k – загальна чисельність населення k -го регіону.

Наведений перелік показників бідності, для оцінки яких доцільно застосувати методи непрямого оцінювання, визначено спеціально створеною робочою групою, яка включала провідних експертів із питань моніторингу бідності.

Рівень оцінювання

Підвищення рівня надійності визначених показників бідності забезпечується для регіонів України. Водночас на основі уточнених оцінок показників для регіонів розраховуються уточнені оцінки показників для рівня економічних районів та України в цілому. Це пояснюється доцільністю забезпечення узгодженості оцінок показників на різних рівнях агрегації даних.

Методи оцінювання показників

Для непрямого оцінювання показників бідності на рівні регіонів України застосовано композиційну оціночну функцію, яка визначається за формулою:

$$P_{comp,k}^{(j)} = \hat{\gamma}_k P_{dir,k}^{(j)} + (1 - \hat{\gamma}_k) P_{synt,k}^{(j)}, \quad j = 1, 2, 3, \quad (4)$$

де $P_{comp,k}^{(j)}$ – оцінка j -го показника бідності для k -го регіону за композиційною оціночною функцією; $P_{dir,k}^{(j)}$ – пряма оцінка j -го показника бідності для k -го регіону; $P_{synt,k}^{(j)}$ – синтетична оцінка j -го показника бідності для k -го регіону; $\hat{\gamma}_k$ – ваговий коефіцієнт композиційної оціночної функції для k -го регіону; індекс j відносить показники рівня бідності до відповідних критеріїв: $j = 1$ – «Рівень бідності за національною межею»; $j = 2$ – «Рівень бідності за регіональною межею»; $j = 3$ – «Рівень бідності за розміром прожиткового мінімуму».

Синтетичні оцінки показників бідності визначаються для кожного показника окремо. Для оцінювання рівня бідності за національною межею побудовано синтетичну оціночну функцію у вигляді наступної регресійної моделі:

$$P_{syn,k}^{(1)} = \beta_0 + \beta_1 x_{1,k} + \beta_2 x_{2,k}, \quad (5)$$

де $\beta_0, \beta_1, \beta_2$ – коефіцієнти регресії, однакові для всіх регіонів України; $x_{1,k}$ – рівень бідності за національною межею у k -му регіоні у попередньому році; $x_{2,k}$ – кінцеві споживчі витрати домогосподарств у розрахунку на одну особу за даними національних рахунків для k -го регіону для поточного року.

Основними характеристиками адекватності регресійної моделі є коефіцієнт детермінації R^2 , F -відношення. Коефіцієнт детермінації R^2 змінюється від 0 до 1 і показує частку змін результативної ознаки (дисперсії результативної ознаки), яка пояснюється регресійною моделлю. Чим ближчі значення R^2 до 1, тим краще лінійна регресійна модель апроксимує результативну ознаку. У загальному випадку критичними для адекватності регресійної моделі слід вважати значення R^2 , які є меншими за 0,7. F -відношення відображає ступінь взаємозв'язку між результативною і факторними ознаками. Якщо розрахункове значення F -відношення перевищує порогове значення $F_{кр}$ (порогове значення

визначається за спеціальними формулами або з довідників), то це свідчить про наявність взаємозв'язку між результативною та факторними ознаками. Таким чином, про адекватність моделі регресії наближено можна зробити висновок, якщо $R^2 \geq 0,7$ та $F > F_{кр}$.

(Коментар: Наведено приклад уточнення оцінки рівня бідності, визначеного за національною межею. Після впровадження моделей буде наведено опис оціночних функцій і для інших показників бідності).

Для забезпечення узгодженості даних на різних рівнях їх агрегації на основі уточнених оцінок показників регіонального рівня розраховуються уточнені оцінки показників для рівня економічних районів та для національного рівня. Слід зауважити, що ці оцінки можуть дещо відрізнятися від прямих оцінок, розрахованих безпосередньо за даними обстеження умов життя домогосподарств.

При визначенні уточненого рівня бідності $P_{comp,U}^{(j)}$ за будь-яким із визначених критеріїв на національному рівні застосовується така формула:

$$P_{comp,U}^{(j)} = \frac{\sum_{k=1}^K N_k P_{comp,k}^{(j)}}{\sum_{k=1}^K N_k}, \quad (6)$$

де N_k – чисельність населення k -го регіону.

Для економічних районів уточнені оцінки показників бідності розраховуються за аналогічними формулами, але лише з урахуванням інформації по тих регіонах, які входять до складу відповідних економічних районів.

Оцінка надійності показників бідності

Основною характеристикою надійності непрямих оцінок показників бідності є величина середньоквадратичної похибки $MSE(P_{comp,k}^{(j)})$, яка визначається з використанням спеціальних статистичних процедур [2]. На основі оцінок величини $MSE(P_{comp,k}^{(j)})$ розраховуються величини відношення кореня квадратного від середньоквадратичної похибки до композиційної оцінки показника бідності за регіонами:

$$RRMSE(P_{comp,k}^{(j)}) = \frac{\sqrt{MSE(P_{comp,k}^{(j)})}}{P_{comp,k}^{(j)}} \cdot 100\%. \quad (7)$$

Величина $RRMSE(P_{comp,k}^{(j)})$ є аналогом коефіцієнтів варіації для прямих оцінок і може бути прийнята за основний критерій надійності результатів оцінювання показників бідності для регіонів України. Непряма оцінка показника бідності може вважатися достатньо надійною для регіонального рівня, якщо значення $RRMSE(P_{comp,k}^{(j)})$ не перевищує 15%. Якщо значення $RRMSE(P_{comp,k}^{(j)})$ знаходиться в інтервалі 15–25%, використання показника рекомендується лише для якісного аналізу. Якщо значення $RRMSE(P_{comp,k}^{(j)})$ перевищує 25%, показник не рекомендується використовувати для аналізу бідності.

Важливою характеристикою надійності оцінювання показників бідності є величина граничної похибки вибірки $ME(P_{comp,k}^{(j)})$ за регіонами, яка для довірчої ймовірності 95% визначається за формулою:

$$ME(P_{comp,k}^{(j)}) = 1.96 \cdot RMSE(P_{comp,k}^{(j)}), \quad (8)$$

де $RMSE(P_{comp,k}^{(j)}) = \sqrt{MSE(P_{comp,k}^{(j)})}$ і є аналогом стандартної похибки. Гранична похибка $ME(P_{comp,k}^{(j)})$ визначає межі довірчих інтервалів композиційних оцінок: $P_{comp,k}^{(j)} \pm ME(P_{comp,k}^{(j)})$.

Технічні зауваження

Всі розрахунки за наведеними моделями здійснюються за затвердженою методологією з використанням спеціального програмного забезпечення, розробленого у середовищі SPSS.

Користування результатами

Статистичні дані та методологічні матеріали, наведені в публікації, призначені для використання органами державного управління, які здійснюють моніторинг бідності в Україні, фахівцями науково-дослідних установ та вищих навчальних закладів, які здійснюють дослідження рівня життя населення України.

Додаткова інформація

Для отримання додаткової інформації щодо результатів обстеження, відповідного методико-методологічного забезпечення, а також довідок щодо умов поширення результатів ОУЖД звертатися:

адреса: 01023, м. Київ, МСП, вул. Шота Руставелі, 3

телефони: (044) 234-01-34, 235-31-22, 234-32-13, 287-12-11

факс: (044) 235-37-39

електронна пошта: households@ukrstat.gov.ua;

I.Osipova@ukrstat.gov.ua.

Основні терміни та поняття

<i>Допоміжна (зовнішня) інформація</i>	– додаткова інформація до оцінок змінних за регіонами
<i>Композиційна оціночна функція</i>	– оціночна функція отримана на основі композиції (комбінації) різних оціночних функцій. В офіційній статистиці частіше за все використовується композиція оціночних функцій, побудованих для прямого та непрямого оцінювання
<i>Непряме оцінювання</i>	– оцінювання показника на основі непрямой оціночної функції, при побудові якої передбачається використання даних, отриманих не лише за результатами безпосереднього вимірювання (на основі конкретного обстеження), а й з використанням інформації з додаткових джерел (наприклад, даних аналогічних обстежень по інших територіях, за інші періоди часу, дані інших обстежень тощо)
<i>Оцінка показника</i>	– значення показника, яке отримано за неповними даними щодо сукупності, яка досліджується. Може бути отримана шляхом розповсюдження даних вибіркової сукупності на генеральну, застосування моделей, визначення значень експертами тощо
<i>Оціночна функція</i>	– правило або співвідношення (формула) для оцінювання показника з використанням даних обстеження. Оціночна функція відома ще до отримання результатів обстеження. Слід розрізняти поняття оціночна функція та оцінка, так як оціночна функція – це функція вибірки, тоді як оцінка є її значенням, знайденим на основі реалізації вибірки
<i>Пряме оцінювання</i>	– оцінювання показника на основі прямої оціночної функції, при побудові якої передбачається, що оцінювання здійснюється на основі даних, отриманих за результатами безпосереднього вимірювання (на основі конкретного обстеження). Допускається також використання даних із додаткових джерел, але такі дані повинні відповідати сукупності, для якої оцінюються показники
<i>Середньоквадратична похибка</i>	– характеристика надійності оцінки показника, яка визначається як сума дисперсії та квадрату величини зміщення

Перелік посилань

1. Rao J.N.K. Small Area Estimation / John Wiley&Sons, 2003. – 314 p.
2. Enhancing Small Area Estimation Techniques to meet European Needs (EURAREA) // Project IST-2000-26290. The Final Report. 2004.

Перелік використаних джерел

1. Kish L. Survey sampling.– Wiley Classics Library Edition Published, 1995. – 643 p.
2. Кокрен У. Методы выборочного исследования. – М.: Статистика, 1976. – 440 с.
3. Ghosh M., Rao J. N. K. Small Area Estimation: An Appraisal, Statistical Science, 1994, Vol. 9, No1, p. 55–93.
4. Rao J.N.K. Small Area Estimation / John Wiley&Sons, 2003. – 314 p.
5. Longford N. T. Missing Data and Small-Area Estimation. Modern Analytical Equipment for the Survey Statistician. Springer-Verlag, New York, 2005.
6. Longford N.T. Estimation of the poverty rates in the regions of Ukraine. – SNTL and UPF, Barcelona, Spain. – The report prepared for the Social Assistance System Modernization Project, Ukraine.
7. Enhancing Small Area Estimation Techniques to meet European Needs (EURAREA) // Project IST-2000-26290. The Final Report. 2004.
8. Додаток 1 до другого звіту за договором № 4807-56 від 22 жовтня 2008 року щодо надання консультативних послуг з моніторингу та оцінки соціальної політики між Міністерством праці та соціальної політики України та компанією ECORYS Netherland B.V., у рамках реалізації спільного з Міжнародним банком реконструкції та розвитку проекту "Удосконалення системи соціальної допомоги": "Пропозиції щодо методологічних підходів до розробки статистико-математичних моделей з метою підвищення якості вимірювання основних показників бідності на регіональному рівні".
9. Методичні рекомендації щодо оптимізації плану вибірки обстеження умов життя домогосподарств з метою підвищення надійності оцінки показників бідності на регіональному рівні. – К: Держкомстат, 2010.
10. Додаток 2 до другого звіту за договором № 4807-56 від 22 жовтня 2008 року щодо надання консультативних послуг з моніторингу та оцінки соціальної політики між Міністерством праці та соціальної політики України та компанією ECORYS Netherland B.V., у рамках реалізації спільного з Міжнародним банком реконструкції та розвитку проекту "Удосконалення системи соціальної допомоги": "Аналітичні матеріали щодо інформаційних джерел, які можуть бути використані при розробці моделей для оцінювання бідності, та їх подальшого практичного застосування".
11. Афифи А., Эйзен С. Статистический анализ: Подход с использованием ЭВМ. – М.: Мир, 1982. – 488 с.
12. Дрейпер Н., Смит Г. Прикладной регрессионный анализ (в 2-х томах). – М.: Финансы и статистика, 1986.

13. Додаток 5 до сьомого звіту за договором № 4807-56 від 22 жовтня 2008 року щодо надання консультативних послуг з моніторингу та оцінки соціальної політики між Міністерством праці та соціальної політики України та компанією ECORYS Netherland B.V., у рамках реалізації спільного з Міжнародним банком реконструкції та розвитку проекту "Удосконалення системи соціальної допомоги": "Пропозиції щодо побудови прототипів статистико-математичних моделей для оцінювання бідності, які забезпечуватимуть належні рівні надійності визначених показників бідності на регіональному рівні та здійснити апробацію прототипів моделей".

14. Додаток 7 до сьомого звіту за договором № 4807-56 від 22 жовтня 2008 року щодо надання консультативних послуг з моніторингу та оцінки соціальної політики між Міністерством праці та соціальної політики України та компанією ECORYS Netherland B.V., у рамках реалізації спільного з Міжнародним банком реконструкції та розвитку проекту "Удосконалення системи соціальної допомоги": "Макет вихідної інформації статистико-математичних моделей для оцінювання бідності, які забезпечують належні рівні надійності визначених показників бідності на регіональному рівні".