Proceedings of CONF-SPML 2026 Symposium: The 2nd Neural Computing and Applications Workshop 2025
DOI: 10.54254/2755-2721/2026.TJ30959

Dynamic Pilot Optimization of South Korea's Urban-Rural Fertility Policies Based on Improved Sliding Window UCB Algorithm

Yang Gou

School of Automation, Hangzhou Dianzi University, Hangzhou, China
g18808110033@outlook.com

Abstract. South Korea is facing a significant low-fertility rate issue, with varying success in fertility policy outcomes between urban and rural areas. The traditional fixed-area pilot model struggles to adapt to non-stationary fluctuations in fertility rates, leading to high trial-and-error costs. This study addresses the optimization of urban-rural fertility policies by proposing an enhanced Sliding Window Upper Confidence Bound (SW-UCB) algorithm that combines a sliding window with a forgetting factor. It treats Seoul and South Jeolla Province as arms of a Multi-Armed Bandit model, defining the increase in fertility rate per unit subsidy as the reward and conducting a simulated 60-month pilot. The improved algorithm demonstrates a 22.3% reduction in cumulative regret compared to the traditional UCB algorithm and a 19.4% reduction compared to Thompson Sampling, effectively accommodating fluctuations in fertility rates and aiding the precise adjustment of policies.

Keywords: Urban-Rural Fertility Policy in South Korea, Multi-Armed Bandit, Sliding Window UCB Algorithm, Non-Stationary Optimization

1. Introduction

Since 2023, South Korea has been confronted with a persistently low fertility rate, and the implementation effects of fertility policies vary significantly between urban and rural areas. In urban areas exemplified by Seoul, childcare subsidy policies have had a minimal impact on augmenting the fertility rate due to high living costs. In contrast, rural areas such as South Jeolla Provinceexhibit sharp fertility rate fluctuations influenced by busy farming seasons. This situation has created a quandary for the traditional fixed - area pilot model, characterized by high trial-and-error costs and low efficiency in policy optimization. Existing studies have provided partial solutions. Through theoretical analysis and experimental verification, Garivier and Moulines introduced the Sliding Window Upper Confidence Bound (SW-UCB) algorithm, theoretically and experimentally verifying that the sliding window technique can effectively capture the abrupt change characteristics of reward distribution in non-stationary environments, serving as a core framework for decision optimization in such scenarios [1]. Building on this, Fiandri et al. further expanded the research by proposing the Sliding Window Thompson Sampling algorithm, systematically validating the dynamic tracking capability of the sliding window technique in non-stationary environments and confirmed the

DOI: 10.54254/2755-2721/2026.TJ30959

effectiveness of combining sliding window with forgetting factor in addressing the Multi-Armed Bandit (MAB) problem [2]. This study addresses two main issues: developing a multi-armed bandit (MAB) model for South Korea's urban-rural diversity and enhancing the UCB algorithm to address fluctuating fertility rates. It introduces an improved SW-UCB algorithm that utilizes a sliding window and forgetting factor, validated through a 60-month simulation, aiming to offer a data-driven tool for accurate policy adjustments.

2. Related theories and model foundations

2.1. Core theory of Multi-Armed Bandit (MAB)

2.1.1. MAB framework

Arm: Corresponding to optional policy schemes. In this study, two "arms" are set up: the Seoul Arm and the South Jeolla Province Arm, representing differentiated policy options for urban and rural areas.

Reward: Measuring policy effectiveness, defined as the "fertility rate increase value per unit subsidy".

Time Step: It refers to the iteration cycle of decision-making. Aligned with the monthly statistical nature of fertility data, the total number of time steps matches the 60-month cycle of the simulated pilot.

2.1.2. Principle of classic UCB

The classic Upper Confidence Bound (UCB) algorithm addresses the "exploration-exploitation" trade-off inherent in MAB by integrating "known returns" and "unknown risks" [3]. Its decision-making formula is as follows:

$$(UCB(a_i) = \overline{x_{a_i}} + c\sqrt{\frac{2lnt}{n_{a_i}}}) \tag{1}$$

The meaning of each parameter in the formula and its adaptability to policy scenarios are as follows:

$\left(\overline{x}_{a_i}\right)$: The historical average reward of the (a_i) -th "arm" (e.g., the fertility rate increase value per unit subsidy of the South Jeolla Province Arm), representing the exploitation dimension, prioritizing the selection of policies that have proven effective.

t: The current total time step (e.g., $(t = 4)$ for the 4th month of the pilot). (n_{a_i}) : The number of times the (a_i) -th "arm" has been selected (e.g., the Seoul Arm has been piloted twice). The smaller the denominator (n_{a_i}) , the larger the second term, which forces the algorithm to "explore" policies with fewer pilot times and avoids missing potential optimal options.

c: Exploration coefficient controlling the intensity of exploration. An excessively large c may lead to a sharp increase in trial-and-error costs, while an excessively small c may result in being trapped in "local optimality".

DOI: 10.54254/2755-2721/2026.TJ30959

2.1.3. Regret

Regret, a core MAB evaluation indicator, quantifies the gap between actual decision - making returns and theoretical optimal returns. Its definition and the characteristics of the classic UCB are as follows:

Definition: It is divided into instantaneous Regret and cumulative Regret.

Instantaneous Regret refers to the difference between the reward of the selected "arm" and that of the optimal arm in a single time step (the policy with the best theoretical effect) within a single time step. For example, if the reward of the Seoul Arm is 0.12 and the optimal arm is the South Jeolla Province Arm with a reward of 0.15 in a certain time step, the instantaneous Regret is 0.03.

Cumulative Regret is the sum of instantaneous Regret over the total time steps T (expressed as $\left(R_T = \sum_{t=1}^{T} (\max \mu - \mu_{a_t})\right)$, where (μ) is the true reward expectation of the "arm"). It represents the degree of resource waste throughout the pilot process.

2.2. The key technology of the non - stationary MAB algorithm

2.2.1. Core formula

Core formula (SW - UCB): Based on the classic UCB, the decision - making index is computed exclusively based on data within the sliding window, and the formula is modified as:

$$(SW - UCB(a_i) = \overline{x_{a_i}^{W}} + c\sqrt{\frac{2lnt}{n_{a_i}^{W}}}) \tag{2}$$

where $\left(\overline{x}_{a_i}^{W}\right)$ is the average reward of the (a_i) - th "arm" within the window, and $\left(n_{a_i}^{W}\right)$ is the number of times the (a_i) - "arm" is selected within the window. When the decision - making step $(t > W)$, the window slides automatically with time, enabling real-time tracking of fertility rate fluctuations.

2.2.2. SW-UCB algorithm design and model construction

Multi-Armed Bandit algorithms in non-stationary environments offer foundational technical support for the dynamic optimization of fertility policies. Among these algorithms, the Sliding Window Upper Confidence Bound (SW-UCB) algorithm proposed by Garivier and Moulines forms a key foundation. Building on this, this study takes "region-policy adaptability" as the core principle and incorporates the disparities in the implementation of urban-rural fertility policies in South Korea to construct an improved SW-UCB model. This model embodies a design logic of "technical reference + scenario innovation":

This study employs the core sliding window mechanism of the SW-UCB algorithm. Considering the 2-month short-term fluctuation cycle of South Korea’s fertility rate—key nodes such as busy farming seasons and graduation seasons occur at intervals of approximately 2 months—a 2-month sliding window is set for busy farming seasons (W=2). This window centers on recent busy farming season data to discern Seoul’s stable advantages while excluding interference from non-busy farming seasons. For non-busy farming seasons, a 3-month sliding window is set (W=3), which uses

DOI: 10.54254/2755-2721/2026.TJ30959

sufficient recent data to validate the high-efficiency characteristics of rural areas and avoid misjudgments caused by insufficient samples.

Meanwhile, a forgetting factor is incorporated. The forgetting factor is set to 0.9 (slow forgetting) for busy farming seasons to to preserve the historical patterns of Seoul's advantages. For non-busy farming seasons, the forgetting factor is set to 0.4 (fast forgetting) to mitigate the impact of outdated data on temporary fluctuations and prioritize decision-making based on the latest high-yield data from South Jeolla Province. By assigning higher decision-making influence to recent data through exponentially decaying weights, the algorithm balances "fluctuation sensitivity" and "data stability," addressing the issue of single-month outlier interference that may be caused by the "rigid truncation" of historical data in the traditional UCB algorithm. This design aligns with the logic of "discount weight optimization for non-stationary decision-making" proposed by Raj and Kalyani [4].

In response to urban-rural differences in South Korea, the "arms" are specifically defined into two types:

Seoul Arm: The policy offers a monthly subsidy of 1 million KRW for families with children aged 0–1, and 500,000 KRW for children aged 1–2. Each newborn is eligible for a total subsidy of 29.6 million to 30.6 million KRW by the age of 8 [5].

South Jeolla Province Arm: The policy offers a monthly subsidy of 200,000 KRW until the child turns 18, resulting in a total subsidy of 43.2 million KRW [6].

This more realistic effect feedback enables the improved SW-UCB algorithm to more accurately determine "when to allocate more funds to Seoul and when to allocate more funds to South Jeolla Province".

3. Simulation experiment and result analysis

3.1. Experiment design

3.1.1. Experimental objectives and subjects

The core objective of the experiment is to validate the adaptability of the improved Sliding Window UCB algorithm to the seasonal fluctuations of rural fertility policies in South Korea, with the traditional UCB algorithm and Thompson Sampling algorithm as benchmarks. The research subjects are set as Seoul (urban policy benchmark, with sufficient childcare resources and stable effects) and South Jeolla Province (rural policy sample, with a high proportion of agricultural population and significant fluctuations in policy effects during busy farming seasons). The simulation period is set to 60 months (5 years), encompassing 5 complete busy/non-busy farming cycles to validate the algorithm's long-term stability.

3.1.2. Parameter settings

Simulation period = 60 months: It covers 5 complete busy/non-busy farming cycles, avoiding interference from short-term fluctuations and verifying the long-term stability of the algorithm.

Seoul policy effect coefficient = 0.26: This urban benchmark value is derived from existing literature [7]. The number of births in 2024 increased by 1,901 year-on-year over a 7-month period, with a growth rate of 8.4%. According to, each newborn can receive a total subsidy of 29.6 million to 30.6 million KRW before the age of 8, averaging 312,500 KRW per month [5]. Based on this, the fertility rate increase per unit subsidy is computed as 0.26.

DOI: 10.54254/2755-2721/2026.TJ30959

Derived from existing literature [6], the number of births increased by 2,231 from January to March 2025, representing a year-on-year growth rate of 6.5%. With a monthly subsidy of 200,000 KRW until the child turns 18 (total 43.2 million KRW), corresponding to an average monthly subsidy of 200,000 KRW, the fertility rate increase per unit subsidy is computed as 0.31. Rural busy season effect coefficient is estimated at 0.20, with a fertility rate of 1.03 for South Jeolla Province in 2024 [8]. During busy farming seasons (May-June, September-October), the fertility rate is 0.8, representing 70% of the non-busy season rate of 1.07. The sliding window size is 2 months for busy seasons and 3 months for non-busy seasons, balancing timeliness and stability. A forgetting factor of 0.9 is applied during busy seasons, and 0.4 during non-busy seasons, to minimize historical data interference. The exploration coefficient is 0.3 for busy and 2.5 for non-busy seasons to optimize decision-making in rural areas [8].

3.1.3. Explanation of parameter changes

According to [8], in dynamic environments, the core of algorithm improvement resides in parameter tuning. Whether it is incorporating an environmental change rate into UCB or adjusting the posterior distribution for Thompson Sampling (TS), the essence is to make the algorithm adapt to environmental changes [9]. In this study, the sliding window size, forgetting factor, and exploration coefficient all change with the shift between busy and non-busy farming seasons. This adaptive design enhances the SW-UCB algorithm's ability to adapt to abrupt environmental shifts.

3.2. Experimental results and analysis

In the 60-month fertility policy simulation, the improved SW-UCB algorithm exhibited significant advantages compared to the traditional UCB and Thompson Sampling (TS) algorithms, realizing comprehensive improvements in both resource utilization efficiency and decision-making precision.

The total cumulative Regret of the improved SW-UCB was 1.500, marking a significant breakthrough relative to the two benchmark algorithms:

Compared with the traditional UCB (1.500 vs. 1.930), it achieved a 22.3% reduction in excessive resource waste across the 5-year period.

Compared with Thompson Sampling (1.500 vs. 1.860), it enhanced resource utilization efficiency by 19.4%.

During the 20 busy farming months, the improved SW-UCB maintained a 50.0% correct selection rate, showing clear advantages over competing algorithms:

It was 15 percentage points higher than the traditional UCB, mitigating decision biases arising from over-reliance on the overall historical mean.

It was 5 percentage points higher than Thompson Sampling, reducing selection variability caused by random sampling.

The core reason resides in the algorithm's seasonally adaptive design: targeted parameter combinations are employed during busy farming seasons to quickly filter out irrelevant data interference and accurately capture Seoul's advantageous performance in such periods. In contrast, the traditional UCB constrained by the historical data mean, and Thompson Sampling is susceptible to randomness, rendering both less capable of achieving stable optimal decision-making.

DOI: 10.54254/2755-2721/2026.TJ30959

4. Rationality of SW-UCB algorithm model settings and adaptation to practical scenarios

From the perspective of practical policy implementation logic, although the setting of "allocating resources to only one region (Seoul/South Jeolla Province) at a time" simplifies the decision-making dimension, the essence of single-region selection in the algorithm constitutes a simplified abstraction of the practical constraint of limited policy resources—not a denial of the feasibility of multi-regional investment. The core logic includes two points: The document discusses the importance of decision-making in resource allocation, emphasizing the need to determine a "priority tilt ratio" for investments across different regions. It outlines how algorithms can shift from "ratio decision-making" to "priority decision-making" by assessing unit resource efficiency, favoring areas like rural regions while still allocating some resources to cities like Seoul. The document also mentions the fair Multi-Armed Bandit (FMAB) algorithm [9], which aims to minimize regret while ensuring all options have selection opportunities. Additionally, it highlights how a simplified model can effectively demonstrate decision-making differences between algorithms, validating the improved capabilities of SW-UCB in identifying seasonal patterns for future complex application [10].

5. Conclusion

This study centers on the research question of "how to enhance the adaptability of algorithms to the seasonal fluctuations of rural fertility policies in South Korea". The analysis results based on 60-month policy simulation data show that: through the differentiated parameter design—2-month sliding window + 0.9 slow forgetting factor for busy farming seasons, and 3-month sliding window + 0.4 fast forgetting factor for non-busy farming seasons—the improved SW-UCB algorithm effectively overcomes the decision-making lag constraint of the traditional UCB algorithm (caused by over-reliance on full historical data) and mitigates the selection fluctuation problem of Thompson Sampling (resulting from randomness). From the perspective of quantitative results, the total cumulative Regret of the improved SW-UCB algorithm is 22.3% lower than that of the traditional UCB and 19.4% lower than that of Thompson Sampling. In the busy farming season scenario, its correct selection rate is 50.0%, which is 15 percentage points higher than the traditional UCB and 5 percentage points higher than Thompson Sampling. The study validates the improved SW-UCB algorithm's adaptability to seasonal fluctuations in fertility policies, addressing the research question regarding optimization algorithms for rural fertility policies. It presents a viable solution for dynamic policy resource allocation. Limitations include restricted regional samples and the exclusion of external variables like subsidy intensity and aging. Future research should broaden the sample to agricultural regions, improve algorithm applicability with field data, and explore policy portfolio optimization and national-level resource allocation.

References

[1] Garivier, A., Moulines, E. (2011). On upper-confidence bound policies for non-stationary bandit problems [C]. pp. 127-152. https: //doi.org/10.1007/978-3-642-24412-4_16
[2] Fiandri, M., Metelli, A.M., Trovò, F. (2025). Sliding-Window Thompson Sampling for Non-Stationary Settings. arXiv: 2409.05181 [stat.ML]. https: //arxiv.org/pdf/2409.05181
[3] setoffai123.my. (2025, March 7). Detailed Explanation of Multi-Armed Bandit Problem Based on Upper Confidence Bound (UCB) Algorithm. https: //setoffai123.my/articles/216010100766.php
[4] Raj, A., Kalyani, S. (2017). Discounted Thompson Sampling for Non-Stationary Bandits [J]. https: //arxiv.org/pdf/1707.09727v1.pdf

DOI: 10.54254/2755-2721/2026.TJ30959

[5] Sohu News. (2025, August 1). South Korea's birth rate rebounds: What did the government do right? Sohu. https: //m.sohu.com/a/919603983_656058
[6] Hanguo Huaren Shequ (Korean Chinese Community). (2025, June 2). South Jeolla Province, South Korea: Total fertility rate reaches 1.13 in Q1, returning to normal levels. 360doc. http: //www.360doc.com/content/25/0602/15/75678705_1154592473.shtml
[7] Yao, Q.L. (2025, April 3). Behind South Korea's "reverse" in birth rate. News.cn. http: //www.news.cn/globe/20250331/26f09de779434138956a2336755019fa/c.html
[8] Yu, J. (2025). Application of Dynamic Multi-Armed Bandit in Real-Time System: Comparative Study of Thompson Sampling and Ucb Algorithm [J]. ITM Web Conf., 78, 01029. https: //doi.org/10.1051/itmconf/20257801029
[9] Patil, V., Ghalme, G., Nair, V., Narahari, Y. (2020). Achieving fairness in the stochastic multi-armed bandit problem [J]. Proceedings of the AAAI Conference on Artificial Intelligence, , 34(04), 5379–5386. https: //doi.org/10.1609/aaai.v34i04.5986
[10] Kalyanakrishnan, S., Tewari, A., Stone, P., et al. (2012). PAC subset selection in stochastic multi-armed bandits [C]. In: Proc. Int. Conf. Mach. Learn. (ICML), Edinburgh, UK. pp. 151–158. http: //cs.utexas.edu/~ai-lab/pubs/ICML12-shivaram.pdf