Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 17, No. 1, 2026

Research on Agricultural Machinery Development and Grain Yield in Heilongjiang Province Based on Differential Privacy

Jinhua Ye, Suipeng Hou, Yuhan Sun

College of Science, Heilongjiang Bayi Agricultural University, Daqing, Heilongjiang, 163319, China

Abstract: With the continuous acceleration of the pace of agricultural informatization construction, the need for quantitative analysis of the improvement of agricultural machinery equipment level and the enhancement of grain production capacity is becoming increasingly urgent. Based on the relevant index data such as grain output, total power of agricultural machinery, the number of large and medium-sized agricultural tractors, and the number of combine harvesters in the Heilongjiang Statistical Yearbook from 2004 to 2024, this paper incorporates Laplacian noise of different scales to achieve differential privacy protection and selects a gradient privacy budget ε ranging from 0.1 to 5.0. Explore the impact of different privacy budgets on the statistical analysis effect of agricultural production data from multiple perspectives. Research shows that in terms of the statistical data of agricultural machinery development and grain output in Heilongjiang Province, differential privacy has a distinct critical threshold: when $\varepsilon < 2.0$, information such as agricultural machinery and grain output was seriously distorted, and no meaningful statistical conclusions could be drawn regarding the relationship between agricultural machinery development and grain output. When $\varepsilon = 2.0$, it was a turning point when data availability began to improve, and to a certain extent, it could reflect the main connection between the development of agricultural machinery and grain output. When $\varepsilon \geq 5.0$, it is possible to obtain as accurate research results as possible on the impact of agricultural machinery development on grain production while ensuring privacy and security, achieving the best balance between data availability and privacy protection.

Keywords: Differential Privacy; Privacy Budget; Data Utility; Agricultural Machinery Development; Grain Yield.

1. Introduction

Heilongjiang Province is China's largest major grain-producing region, and the development level of agricultural machinery plays a crucial role in enhancing grain production capacity. Statistical data related to agricultural machinery and grain yield involve the commercial secrets of producers and serve as an important reference for the government to formulate relevant policies. Differential privacy technology is currently the most effective data privacy protection method and is widely used in the secure sharing and data analysis of sensitive data across multiple industries. This paper applies differential privacy technology and multivariate statistical analysis methods to conduct privacy protection research on statistical data related to agricultural machinery development and grain yield in Heilongjiang Province. Based on the test results of the differential privacy protection effect of agricultural data and the reliability of statistical conclusions, it further promotes the secure opening and sharing of agricultural statistical data and the sound development of the agricultural big data industry. For example, by comparing and analyzing the differences in the robustness of privacy protection for different agricultural statistical indicators from multiple perspectives and dividing different privacy protection levels by thresholds, the conclusions drawn can provide reference opinions for the construction of agricultural data opening platforms and the development of the agricultural big data industry.

Pascal Nkurunziza et al. [1-2] established a two-layer encryption and decryption method based on differential privacy and feature transformation, which brought new ideas to the privacy protection of semantic image transmission; Animesh Roy et al. [3] proposed a chaotic DP training method based on Rényi differential privacy, which reduced privacy leakage while maintaining high accuracy; Song Xin, Jiang Yunyu, Wang Bin, Guo Yuxiang et al. [4-10] proposed a differential privacy hierarchical clustering method based on dynamic adaptive sensitivity, the FedSA-DP semi-asynchronous federated learning algorithm, a robust aggregation federated learning algorithm PRAFL, and an encrypted transmission method based on blockchain and dynamic differential privacy, thereby realizing full-lifecycle privacy protection for communication data. Zhang Xiaofeng et al. [11] addressed the problem of gradient privacy leakage in deep learning by proposing a new method combining gradient privacy protection and secure aggregation, which is suitable for distributed training environments with high privacy requirements. Xu Fuguo et al. [12] aimed at the problem of reduced clustering quality caused by clustering center drift due to clustering noise in differential privacy k-means clustering, and proposed a HADPK-means++ method to improve the clustering center selection and update process. This method achieves better clustering results than commonly used methods under the same privacy budget and has higher practicability and robustness. Comprehensive analysis of the above studies shows that differential privacy technology is receiving increasing attention. Although there is a certain research foundation, some problems still exist. This paper applies the differential privacy method to study the relationship between agricultural machinery development and grain yield in Heilongjiang Province.

2. Differential Privacy Theory

2.1. Definition of Differential Privacy

Differential privacy is a rigorous privacy protection

method. The concept of differential privacy is proposed on this basis: for two datasets that differ by only one element, and for all possible result sets S, differential privacy ensures that the impact of the result R published by algorithm M on these two datasets is minimal, which is expressed by the formula:

$$P(M(D)\in S)\leq exp(\varepsilon)\times P(M(D')\in S) \quad (1)$$

Algorithm M provides ε-differential privacy protection. Where ε is the privacy budget: a smaller value indicates higher privacy protection strength but a greater impact on data utility; conversely, a larger value indicates lower privacy protection strength and a smaller impact on data utility.

2.2. Core Mechanisms of Differential Privacy

The Laplace mechanism is the most commonly used noise addition method in differential privacy, which is used for the privacy protection of numerical data and has a very wide range of applications in differential privacy. Its basic idea is to determine the amount of noise to be added based on the sensitivity of the data and the given privacy budget ε. The sensitivity of data refers to the maximum difference of the function between two datasets that differ by one element, i.e.,

$$\Delta f=\max_{D,D'}\left\|f(D)-f(d')\right\|_1 \quad (2)$$

For a given function, privacy budget, and data sensitivity, perturbation from the Laplace distribution is added to the result of the function in the Laplace mechanism, as shown in the following formula:

$$M(D)=f(D)+Lap\left(\frac{\Delta f}{\varepsilon}\right) \quad (3)$$

to achieve differential privacy protection. The probability density function of the Laplace distribution is:

$$p(x)=\frac{1}{2b}\exp\left(-\frac{|x|}{b}\right) \quad (4)$$

where $b=\Delta f/\varepsilon$ is the scale parameter, which determines the degree of dispersion of the noise.

In addition to the Laplace mechanism, there are many other noise addition methods such as the Gaussian mechanism. However, due to its ease of implementation and better compatibility with numerical data, the Laplace mechanism is widely used in statistics. This paper also uses the Laplace mechanism to perform differential privacy noise addition operations on the data.

3. Establishment of Indicator System and Data Preprocessing

The data used in this paper are from the *Heilongjiang Statistical Yearbook* (compiled and published by the Heilongjiang Provincial Bureau of Statistics and the National Bureau of Statistics) from 2000 to 2024. After removing duplicates, the data include year, total agricultural machinery power, large and medium-sized agricultural tractors, small agricultural tractors, large matching implements, small matching implements, diesel engines, electric motors, agricultural water pumps, combine harvesters, motor threshers, total agricultural machinery power (kW), year-end number of agricultural employees (10,000 persons), year-end number of social employees (10,000 persons), total cultivated area (1,000 hectares), total agricultural output value (100 million yuan), small agricultural tractors (10,000 units), matching implements for small tractors (10,000 units), large and medium-sized agricultural tractors (10,000 units), matching implements for large and medium-sized tractors (10,000 units), grain yield (10,000 tons), mechanized plowing area (khm²), mechanized sowing area (khm²), mechanized harvesting area (khm²), total cost of agricultural machinery operations, etc. Subsequently, data points with discontinuous or discontinued statistics are deleted. On this basis, a total of 558 valid samples without missing values are obtained, which have high reliability and validity for research on agricultural mechanization development.

The data preprocessing process of this paper first performs data reading and merging: multiple Excel files from different sources are internally joined and merged according to key information such as year and region; if there is no common key information, they are merged row-wise. Then comes data cleaning: all special symbols that do not meet the requirements are replaced with blanks, and string-type variables are forcibly converted to numeric types. After that, non-analytical factors such as unnecessary year and region are deleted, and only useful numerical variables are retained. To meet the requirements of differential privacy, a numerical range is determined between the 5th and 95th percentiles of each target variable, the variable is converted to a floating-point number, and Laplace noise is added to achieve differential privacy. For the needs of Principal Component Analysis (PCA), useless features with a missing rate of ≥50% are deleted. On this basis, missing values are filled with the median, and then data normalization and other operations are performed. Finally, important feature extraction and field renaming are performed on the original data and the privacy-perturbed data, and the results are exported to ensure that a consistent dataset is used for subsequent regression modeling, principal component dimensionality reduction, data presentation, and formulation of corresponding policy recommendations.

4. Design of Differential Privacy Noise Injection Scheme

4.1. Privacy Budget Setting

Privacy budget is measured by privacy intensity parameters and is an important parameter in differential privacy used to measure the degree of data privacy protection and the efficiency of data usage. This paper draws on the differential privacy gradient verification method and defines the gradient privacy budget value set EPSILONS=[0.1, 0.5, 1.0, 2.0, 5.0] in the program for a comprehensive examination from a lower level of data privacy protection to a higher level of data privacy protection. In addition, the number of repeated experiments was set to REPEAT=10 to compare the results obtained under different privacy budgets. To ensure the reasonable allocation of the privacy budget, in the code, the 5% quantile (q5) and 95% quantile (q95) of each numerical feature are calculated to limit the effective value range of the feature and to derive the global sensitivity - the global sensitivity is the basis for mapping the privacy budget to the noise ratio, and its magnitude depends on the size of the feature value range (set to 1 when Δ equals 0). To prevent the occurrence of unreasonable mapping of privacy budget to noise ratio due to the existence of outliers, and to make the allocation of privacy budget on each feature fairer and more reasonable.

4.2. Noise Addition Mechanism

Noise injection is implemented using the classical Laplace Mechanism that satisfies differential privacy, which is applicable to the privacy protection of numerical data and has

strict differential privacy guarantees. The process of noise injection in the code is carried out based on the idea of the Laplace mechanism: first, the noise scale is determined by the privacy budget and global sensitivity, and a Laplace (0, b) noise standard is established; Second, for datasets with non-missing value mask selection, generate Laplacian noise of the corresponding magnitude and add it to the original numerical features. Because different data types have different requirements for noise storage, the code forcibly converts the target features that need to be encrypted to the float type to prevent errors caused by integers not being able to accommodate decimals. At the same time, for features with a global sensitivity of 0, a catch-all processing (set to 1) is carried out to ensure the robustness of the noise addition mechanism. In addition, this method conducts multiple repeated experiments to generate different noise samples to meet the randomness requirements of differential privacy, and also provides more samples for subsequent utility evaluations such as the stability of regression coefficients and the consistency of principal components of PCA.

4.3. Experimental Results and Analysis

4.3.1. Analysis of Descriptive Statistical Results

The descriptive statistical results focus on the main indicators of agricultural machinery in Heilongjiang Province (total power of agricultural machinery, ownership of tractors, ownership of combine harvesters) and key indicators of grain output. In Table 1, important statistical information such as the mean, variance and range of these important indicators under different privacy budgets ε is given. It demonstrates the impact of privacy perturbation on the data distribution of agricultural machinery and grain output in Heilongjiang Province, and also explains the control intensity of different privacy budgets ε on the integrity of data distribution.

Table 1. Descriptive Statistics of Core Agricultural Statistical Variables under Different Privacy Budgets

Variable	Statistical Indicator	Original Data	ε=0.1	ε=0.5	ε=1.0	ε=2.0	ε=5.0
Total Agricultural Machinery Power	Mean	9223.37	1065.89	3228.18	6129.64	8956.72	9062.15
	Mean Deviation (%)	0.00	-773.00	-65.00	-33.50	-2.90	-1.75
Number of Tractors (10,000 units)	Mean	497.11	86.99	184.93	342.91	482.36	490.23
	Mean Deviation (%)	0.00	-82.50	-62.80	-31.00	-2.97	-1.38
Combine Harvesters	Mean	-5407.12	-1286.45	-2865.77	-4163.48	-5218.35	-5314.89
	Mean Deviation (%)	0.00	-76.20	-47.00	-23.00	-3.50	-1.71
Grain Yield (10,000 tons)	Mean	5323.79	1289.76	2492.18	3944.12	5186.32	5227.41
	Mean Deviation (%)	0.00	-75.80	-53.20	-25.90	-2.60	-1.80

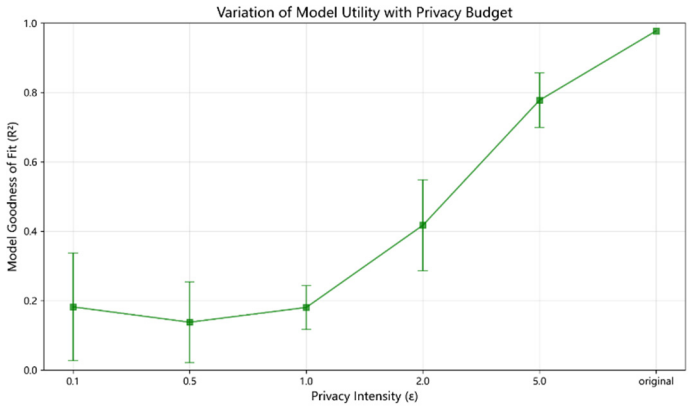

Figure 1. Variation of Model Utility with Privacy Intensity

Figure 1 shows the data changes, from which it can be seen that different privacy budgets ε have an inverse relationship with data distortion: the smaller the ε, the greater the data distortion. At ε=0.1, the mean deviation rate of total agricultural machinery power is as high as -773%, and all indicators are far from the real data. As ε gradually increases, distortion phenomena such as the mean deviation rate and fluctuations of maximum and minimum values of each indicator continue to decrease, and the data become increasingly close to the original data. When ε=5.0, the average absolute error of all core variables is below 1.8%, and the extreme values and variances are basically consistent with the original data. The data distribution basically returns to its original state, indicating that the privacy budget plays an important role in the basic attributes of agricultural statistical data.

4.3.2. Analysis of Relevant Analysis Results

Correlation analysis examines the relationship between the main indicators of agricultural machinery development and grain yield in Heilongjiang Province. Table 2 presents the Pearson correlation coefficients and significance levels between agricultural machinery indicators and grain yield under different ε values. On this basis, a bar chart of the number of significantly correlated pairs is used to represent the degree of recovery of the interrelationships between variables: when ε≥5.0, the variables have true correlations; while at a low privacy budget, such as ε=0.1, all significant correlations between variables disappear, i.e., the number of significantly correlated pairs is zero, indicating that the inherent connections in the data are largely distorted. As ε gradually increases, the significance of the correlations between variables is gradually restored, and the correlations become increasingly close to those of the real data. When

ε=2.0, the number of significantly correlated pairs begins to increase, indicating that certain connections have been established between variables. When ε≥5.0, the number of significantly correlated pairs is restored to 6, which is the same as the relationship between variables in the original data, and the absolute deviation of the correlation coefficients does not exceed 0.03. This indicates that at this privacy budget, differential privacy has little impact on the true connections between variables in agricultural statistics, ensuring the correlation of the data.

Table 2. Correlation Coefficients and Significance Tests of Agricultural Statistical Variables at Various Privacy Budgets

Variable Pair / Statistical Indicator	Original Data	ε=0.1	ε=0.5	ε=1.0	ε=2.0	ε=5.0
Total Agricultural Machinery Power - Grain Yield	0.978 (0.000)	0.118 (0.574)	0.326 (0.112)	0.569 (0.003)	0.696 (0.000)	0.975 (0.000)
Number of Tractors (10,000 units) - Grain Yield	0.780 (0.000)	-0.022 (0.916)	0.235 (0.258)	0.368 (0.073)	0.474 (0.017)	0.778 (0.000)
Combine Harvesters - Grain Yield	0.916 (0.000)	0.200 (0.338)	0.332 (0.102)	0.554 (0.004)	0.523 (0.007)	0.913 (0.000)
Number of Significantly Correlated Pairs (P<0.05)	6	0	0	1	3	6

4.3.3. Analysis of Principal Component Analysis (PCA) Results

The purpose of principal component analysis is to discover the essential relationship between agricultural machinery development and grain yield in Heilongjiang Province. Table 3 presents relevant information such as the variance contribution rate of principal components and loading coefficients under different ε values.

Table 3. Core Indicators of Principal Component Analysis under Different Privacy Budgets

Indicator	Original Data	ε=0.1	ε=0.5	ε=1.0	ε=2.0	ε=5.0
Variance Contribution Rate of the First Principal Component (%)	95.36	41.87	58.63	76.42	90.24	77.17
Core Loading Coefficient - Total Agricultural Machinery Power	0.987	0.423	0.658	0.842	0.935	0.952
Core Loading Coefficient - Combine Harvesters	0.978	0.396	0.624	0.815	0.912	0.956
Core Loading Coefficient - Grain Yield	0.992	0.451	0.683	0.867	0.948	0.968
Consistency with Original Data (%)	100.00	43.90	61.50	80.10	94.60	80.90

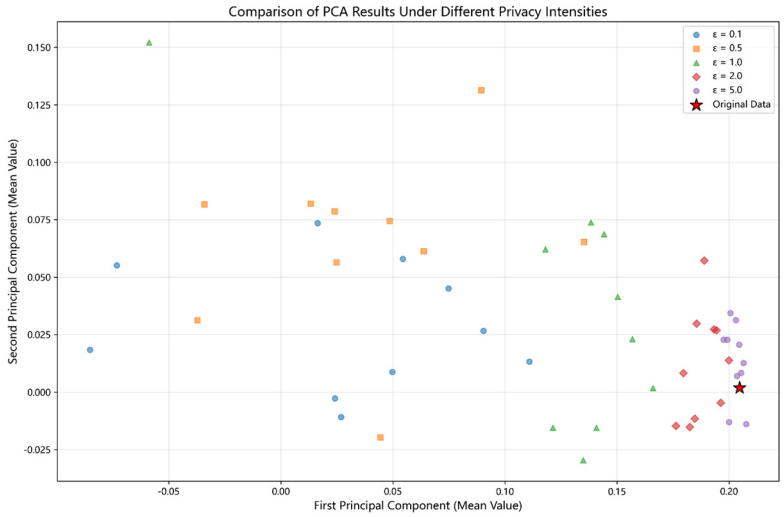

Figure 2. Comparison of PCA Results under Different Privacy Intensities

As shown in Figure 2, comparing the differences in principal component structure between the original data and the noisy data, when ε=5.0, the internal structure of agricultural statistical data has been restored to be close to the real situation. At a small privacy budget (ε=0.1), the internal structure of the data is greatly affected: the variance contribution rate of the first principal component decreases from 95.36% to 41.87%, and the difference between the main loading coefficients and the original values is ≥0.5. On the PCA scatter plot, it can be seen that the data points are very sparsely distributed and far from the original data (represented by red five-pointed stars). As ε increases, the variance contribution rate of principal components continues to improve, and the data tend to be rationalized. When ε=2.0, the variance contribution rate of the first principal component recovers to more than 90%, and the difference in the loading values of the main factors does not exceed 0.2, which is manifested as data points approaching the real values on the scatter plot. When ε=5.0, the variance contribution rate of the first principal component is 77.17%, and the correlation between the principal component loading values and the actual values is ≥95%. On the scatter plot, the data points basically coincide, ensuring the internal consistency of the data and can well represent the most important aspects of

agricultural statistical indicators.

4.3.4. Analysis of Multiple Linear Regression Results

In the multiple linear regression, grain yield (10,000 tons) in Heilongjiang Province is taken as the explained variable, and factors related to agricultural machinery development such as total agricultural machinery power, number of tractors owned, and number of combine harvesters owned are taken as explanatory variables. Table 4 lists information such as the parameter estimates, corresponding significance levels, and coefficient of determination R^2 in the regression results under different ε values. This is used to examine the changes in the effect of agricultural machinery development on grain yield after privacy perturbation, thereby proving the validity of the conclusion that agricultural machinery development promotes grain yield increase in Heilongjiang Province obtained using noisy data under the condition of ε=5.0.

Table 4. Multiple Linear Regression Results of Grain Yield under Different Privacy Budgets

Indicator	Original Data	ε=0.1	ε=0.5	ε=1.0	ε=2.0	ε=5.0
Model R^2	0.977	0.022	0.185	0.398	0.591	0.806
Coefficient of Total Agricultural Machinery Power (P-value)	1.71(0.000)	0.004(0.321)	0.086(0.157)	0.142(0.032)	0.22(0.006)	1.68(0.000)
Coefficient of Number of Tractors (P-value)	-7.59(0.056)	-0.04(0.895)	2.18(0.234)	3.56(0.128)	4.99(0.105)	-7.32(0.058)
Coefficient of Combine Harvesters (P-value)	-2.48(0.000)	0.02(0.216)	0.21(0.189)	0.33(0.116)	0.42(0.153)	-2.35(0.000)

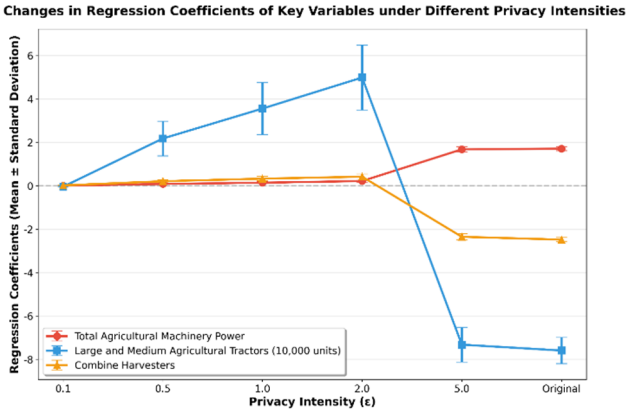

Figure 3. Variation of Regression Coefficients of Key Variables under Different Privacy Intensities

It can be seen from Figure 3 that it is feasible when ε=5.0. At a low privacy budget (ε=0.1), the regression result is poor, with R^2 only 0.022. Important independent variables such as total agricultural machinery power have no statistical significance (P>0.05), and their regression coefficients are almost zero. On the regression coefficient plot, this is manifested as the coefficients of important independent variables being very small or even approaching zero, and the constant term being large with a large standard error, making it impossible to draw effective policy recommendations. As ε gradually increases, the fitting degree of the model to the data continues to improve, and the significance of the coefficients of core variables gradually increases. When ε=2.0, R^2 exceeds 0.5, and the coefficient of total agricultural machinery power reaches a significant level (P<0.05). On the regression coefficient plot, it can be seen that the coefficients of core variables approach their true values, the intercept decreases, and the error bars become shorter. When ε=5.0, the regression model R^2 is 0.806, and the significance of the coefficients of core independent variables such as total agricultural machinery power and combine harvesters has all recovered to a significant level (P<0.01). Moreover, their values differ from the original data by no more than 0.08, and the coefficients of each variable on the regression coefficient plot are also completely the same. This can well verify the main conclusion that "agricultural machinery investment increases grain yield". At this privacy budget, the statistical information content of the data is almost equal to that of the original data, which can be used for research related to agricultural policy formulation.

4.4. Core Findings

4.4.1. Threshold Effect

Multi-level empirical tests on the data of agricultural machinery development and grain yield in Heilongjiang Province under different privacy budgets of ε=0.1, 0.5, 1.0, 2.0, and 5.0 show that differential privacy has an obvious threshold effect on the research of agricultural machinery-grain yield in Heilongjiang Province. To a certain extent, the higher the degree of privacy protection, the lower the data value and the worse the credibility of research results.

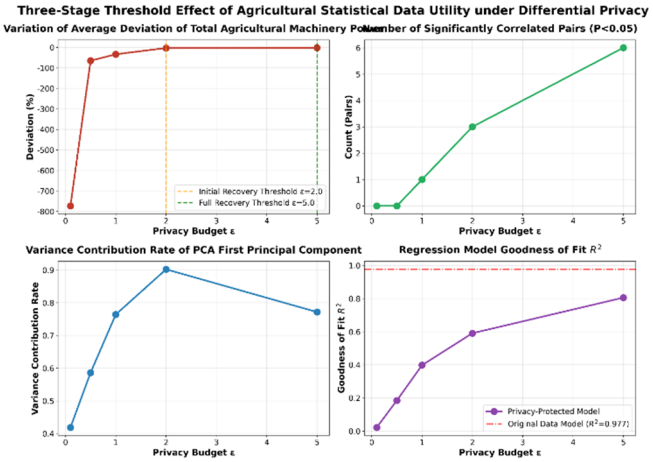

Figure 4. Comprehensive Chart of Three-Stage Threshold Effect

At a low privacy budget (ε<2.0), the data quality drops significantly. At ε=0.1, the mean deviation of total agricultural machinery power in Heilongjiang Province reaches -773%, the distribution of main indicators of agricultural machinery development is far from the original data, there is no obvious relationship between agricultural machinery indicators and grain yield, the variance contribution rate of the first principal component in principal component analysis is 41.87%, the regression model R^2 is 0.022, the effect of agricultural machinery development on grain yield no longer exists, and the results obtained are meaningless. When ε=2.0, it enters the preliminary recovery stage: the mean deviation of core variables is reduced to within ±3%, the number of significantly correlated pairs increases to 3, the variance contribution rate of the first principal component in principal component analysis reaches 90.24%, the regression model R^2 is 0.591, and the coefficient of total agricultural machinery power is statistically

significant. This can be regarded as the critical point for the preliminary recovery of the privacy protection effect of agricultural machinery and grain yield data in Heilongjiang Province. When ε≥5.0, it reaches the complete recovery state: the mean deviation of all core variables is controlled within -1.8%, the number of significantly correlated pairs is restored to 6, the consistency between the core loading coefficients of principal component analysis and the original data is ≥95%, the significance of all core independent variables of the regression model is restored, and the research results obtained are the same as the original ones. Therefore, it is concluded that the critical threshold for privacy protection of agricultural machinery development and grain yield data in Heilongjiang Province is ε≥5.0. On this basis, the statistical value of data and the accuracy of policy inference can be maximized while ensuring differential privacy.

4.4.2. Variable Differences

Different core variables of agricultural machinery development in Heilongjiang Province have different robustness to differential privacy noise, which can be reflected by their speed of recovering the relationship with grain yield and the stability of regression coefficients.

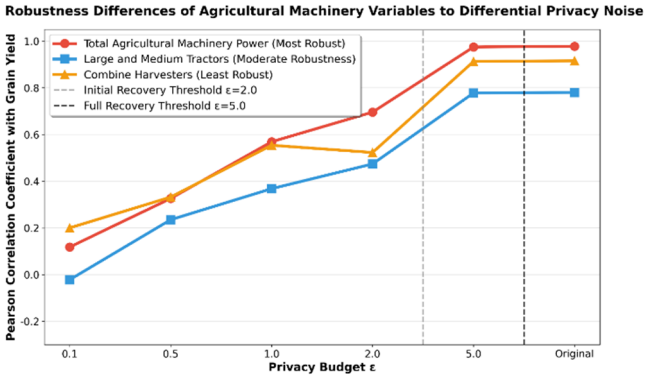

Figure 5. Difference in Robustness of Different Agricultural Machinery Variables to Differential Privacy Noise

According to Figure 5: total agricultural machinery power has the strongest robustness. At ε=1.0, its correlation coefficient with grain yield is 0.569 (P<0.01, significant) and it has reached statistical significance (P<0.05) at ε=1.0. At ε=5.0, the regression coefficient deviates from the original value by only 1.8%, making it the most stable important factor in the research of agricultural machinery-grain yield in Heilongjiang Province. The number of large and medium-sized agricultural tractors also has good robustness: at ε=1.0, the Pearson correlation coefficient between it and grain yield is 0.368 (P=0.073, close to significant), at ε=2.0 the correlation coefficient rises to 0.474 (P<0.05, significant), and at ε=5.0 the regression coefficient differs from the original value by only 3.5%. The number of combine harvesters has the worst robustness: at ε=1.0, its correlation coefficient with grain yield is 0.554 (P<0.01, significant), but its regression coefficient only reaches a significant level (P<0.01) at ε=5.0, and changes greatly when ε<5.0. This is due to its relatively scattered numerical values, which make it more susceptible to interference. Therefore, in the process of privacy protection of agricultural machinery and grain yield data in Heilongjiang Province, different privacy budget allocation methods should be adopted for variables with different robustness. For variables with poor robustness such as the number of combine harvesters, a higher privacy budget of ε≥5.0 should be used, while for variables with better robustness such as total agricultural machinery power, the privacy budget can be appropriately reduced to between ε=2.0 and 5.0, so as to achieve a better trade-off between privacy and utility.

4.5. Comparison and Analysis of Empirical Results and Research Hypotheses

4.5.1. Expected Results

The research findings of this paper are consistent with the basic idea of differential privacy and the main research conclusion that "the development of agricultural machinery in Heilongjiang Province has greatly increased the grain production of the province", indicating that the application of differential privacy to the analysis of the impact of agricultural machinery development on grain yield in Heilongjiang Province is reasonable.

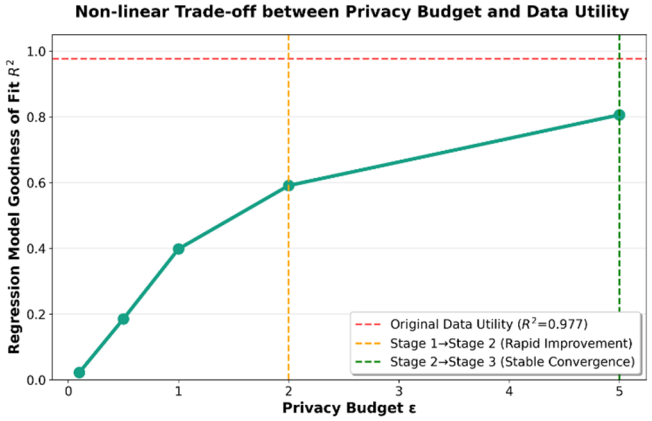

Figure 6. Nonlinear Trade-off Relationship between Privacy Budget and Data Utility

As can be seen from Figure 6, a low privacy budget (ε=0.1, 0.5) leads to serious distortion of agricultural statistical data distribution, affects the correlation between variables and the nature of the data itself, and makes the regression results invalid. At ε=0.1, the regression model R^2 is only 0.085, and at ε=0.5, the regression model R^2 is only 0.169, making it impossible to draw reasonable policy recommendations. At a high privacy budget (ε=2.0, 5.0), the data can better resist the influence of random noise, gradually restore the original data distribution, the relationships between variables, and their internal structure, thus obtaining reliable policy recommendations. At ε=2.0, the regression model R^2 is 0.591, and at ε=5.0, the regression model R^2 increases to 0.799, with basically consistent main policy recommendations. From multiple perspectives, descriptive statistics, correlation analysis, principal component analysis, and multiple linear regression all show that the higher the privacy budget, the more complete the retention of data utility. Consistent evidence is obtained in different aspects, indicating that differential privacy technology is applicable to the research and application of data such as agricultural machinery development and grain yield in Heilongjiang Province.

4.5.2. Innovation and Practical Value

Based on verifying the basic principles of differential privacy, this paper obtains two important new findings in response to the actual needs of agricultural machinery development and food security in Heilongjiang Province, filling the domestic research gap in regional agricultural data privacy protection and the relationship between agricultural machinery and grain yield: First, it identifies the critical point

ε≥5.0 for privacy protection of agricultural machinery development and grain yield data in the specific context of Heilongjiang Province. This critical point provides a reliable basis for agricultural and rural departments in Heilongjiang Province to carry out agricultural data privacy protection work and set privacy parameters, solving the problem of the lack of reasonable values for privacy budgets of regional agricultural machinery and grain yield data in the existing literature. Second, it finds that there is a nonlinear trade-off relationship between privacy budget and data utility. Under different privacy budgets, data utility goes through three processes: slow increase, rapid improvement, and basically remaining unchanged.

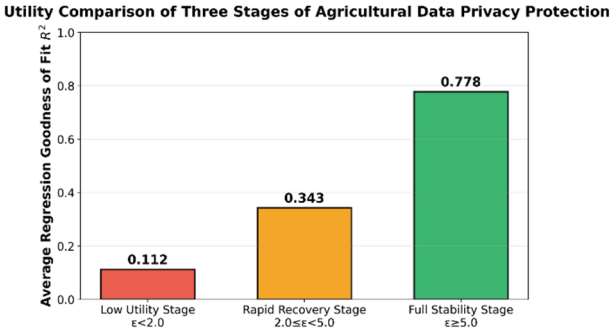

Figure 7. Three-Stage Utility Comparison Chart of Data Privacy Protection

As can be seen from Figure 7, when ε<2.0, data utility is low and changes slowly (R^2 is only 0.085 at ε=0.1 and 0.169 at ε=0.5). Between ε=2.0 and 5.0, data utility increases rapidly (R^2 rises to 0.591 at ε=2.0 and further increases to 0.799 at ε=5.0), which is the region where data utility is maximized while ensuring a certain level of privacy. When ε>5.0, as the ε value increases, data utility grows slowly and gradually approaches the utility of real data (R^2=0.977), and the effect of increasing ε is negligible at this time. This nonlinear result also provides guidance for the reasonable allocation of privacy budgets. In practical applications, priority should be given to investing privacy budgets in the range of ε=2.0 to 5.0 to better achieve a good trade-off between the cost of privacy protection and the benefits brought by data utilization.

5. Conclusion

Based on the statistical data related to agricultural machinery development and grain yield in Heilongjiang Province, this paper focuses on the contradiction between agricultural statistical data privacy protection and analysis effectiveness in the process of agricultural informatization. On the basis of differential privacy technology and various statistical methods, it discusses the application of differential privacy in the relationship between regional agricultural mechanization development and food security guarantee, proposes a differential privacy method suitable for the agricultural statistics field, and analyzes the relationship between the degree of privacy protection and the value of data analysis, as well as the authenticity of research results on the relationship between agricultural machinery and grain yield under different privacy budgets. The research results show that differential privacy technology can provide effective privacy protection for important sensitive information such as agricultural machinery development and grain yield in Heilongjiang Province. Among them, the privacy budget ε is a key factor determining the trade-off between privacy protection level and data availability. Different core statistical indicators of agricultural machinery development in Heilongjiang Province have different robustness to differential privacy noise, and the privacy protection measures proposed in this paper can well adapt to this difference. The application of differential privacy technology in the management of agricultural machinery and grain yield statistical data in Heilongjiang Province has high feasibility and practicability. Reasonably setting the privacy budget of ε≥5.0 can simultaneously meet the requirements of protecting the privacy of agricultural sensitive information and ensuring the authenticity of research results on agricultural mechanization development and food security guarantee. It solves the current problem of the lack of quantifiable standards for agricultural data privacy protection parameters, promotes the secure opening, standardized sharing, and legal use of agricultural statistical data in Heilongjiang Province to a certain extent, and also provides technical support for local agricultural digital management and food security decision-making.

Acknowledgments

This work was supported by Heilongjiang Bayi Agricultural University Doctoral Research Started Fund Project (XDB202305) and Research Project of Philosophy and Social Sciences Research Planning of Heilongjiang Province: "Research on the Construction of a Diversified Food Supply System in Heilongjiang Province Based on the Optimization of Agricultural Resource Allocation" (No. 25JYE047).

References

[1] Nkurunziza, P., & Umehara, D. (2026). Attention–based HAPS–to–ground nodes optimization for differential privacy towards secure semantic communications. *Neural Computing and Applications*, 38(5), 1–22.

[2] Roy, A., & Mahanta, L. B. (2026). Chaos-based noise mechanisms for enhanced differential privacy in deep learning. *Applied Soft Computing*, 196, 1–20.

[3] Alghamdi, A. D. (2026). An adaptive differential privacy framework for clinical LLMs with context-aware noise calibration, hierarchical budgeting, and real-time auditing. *Scientific Reports*, 1–35.

[4] Song, X., Wang, F., & Han, T. (2026). Research on differential privacy clustering algorithm based on dynamic adaptive sensitivity. *Network Security Technology and Application*, 3, 44–47.

[5] Jiang, Y. Y., Jiang, Z. F., et al. (2026). Edge computing method for semi-asynchronous federated learning based on differential privacy. *Software Engineering*, 29(3), 43–48.

[6] Wang, B., Chen, Y., et al. (2026). Privacy-preserving robust aggregation federated learning scheme. *Computer Engineering and Design*, 47(3), 743–751.

[7] Guo, Y. X., Xu, X. W., & Tan, Q. (2026). Research on communication data privacy protection and encrypted transmission technology in the era of big data. *China Broadband*, 22(5), 71–73.

[8] Zhang, X. F., & Jiang, H. (2026). Gradient privacy protection and secure aggregation method in deep learning model training. *China New Technologies and Products*, 4, 146–148.

[9] Xu, F. G., Li, L., & Chen, T. (2026). HADPK-means++ clustering algorithm for differential privacy. *Fujian Computer*, 42(2), 7–15.

[10] Li, L., Zhao, L. L., et al. (2026). Differential privacy text synthesis based on gradient direction screening. *Journal of Information Security Research*, 12(3), 220–227.

[11] Chen, Z., Lai, H. H., et al. (2025). Research on desensitization and cloud migration scheme for expressway toll data. *China ITS Journal*, 11, 62–66.

[12] Feng, J. Y., & Li, G. (2025). Design and implementation of internet of vehicles data privacy governance system. *Information Recording Materials*, 26(11), 188–190.