International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 5286-5295

https://doi.org/10.70517/ijhsa464368

# Numerical Modeling Analysis and Legal Mechanism Research on Personal Information Protection in Cross-border Data Flow

**Zhongguo Lv**[1,*]
[1] Law School, Huainan Normal University, Huainan, Anhui, 232038, China
Corresponding authors: (e-mail: lvzhongguo@edu.cn).

**Abstract** The acceleration of global digital transformation and the widespread application of emerging technologies such as big data, artificial intelligence, and cloud computing have contributed to the exponential growth in the scale of data exchange between countries. This study analyzes the influence mechanism of the level of cross-border data flow on the effect of legal regulation on personal information protection by means of a multiple linear regression model. Based on the panel data of 12 exporting countries and 30 importing countries from 2016-2024, a regression model containing control variables such as the level of economic development, geographic distance, population size, the level of Internet infrastructure development, and the level of foreign direct investment was constructed. The results show that every 1 percentage point increase in the level of cross-border data flow enhances the legal regulation effect of personal information protection by 0.792 percentage points, and the model coefficient of determination improves from 0.603 to 0.914. The endogeneity test shows that every 1% increase in the level of cross-border data flow leads to a 7.9% increase in the legal regulation effect of personal information protection under the treatment of the instrumental variable method. In the robustness test, the impact coefficient of adding the personal information protection awareness variable is 0.176 and significant at the 1% level. The study finds that the level of cross-border data flow promotes the legal regulation effect of personal information protection by influencing the level of national economic development and infrastructure construction, which provides empirical support for improving cross-border data governance.

**Index Terms** cross-border data flow, personal information protection, legal regulation effect, multiple linear regression model, data governance, endogeneity test

## I. Introduction

In today's globalization, cross-border activities are becoming more and more frequent, and cross-border data flow of personal information is becoming more and more common [1], [2]. Whether in shopping, socializing or working, our personal information may be transferred to other countries or regions [3]. For enterprises, cross-border data flow can help them expand overseas markets, for example, by analyzing the data of overseas customers and precisely adjusting product and marketing strategies [4], [5]. For scientific research, cross-border data flow allows scientists to access research data from all over the world and accelerate the output of scientific research results [6]. It can be seen that cross-border data flow has many benefits, however, there are differences in the laws and standards for personal information protection in different countries and regions, which brings risks to the security of personal information [7], [8].

The regulation of cross-border data transfer by personal information protection laws stems from the inevitability of data flow in the context of globalization in the digital age [9], [10]. The legislation of various countries generally regards personal information as a special resource with sovereignty attributes, and China's Personal Information Protection Law explicitly requires that data exiting the country must go through legal procedures such as security assessment, authentication, or signing a standard contract [11]-[13]. This legal basis reflects both the protection of individual rights and the maintenance of national data sovereignty [14]. The legitimacy framework for cross-border data transmission contains three levels: the constraints of international treaties, the regulation of domestic legislation, and the supplementation of industry self-regulation, and the construction of this multi-level legal system safeguards the economic needs of the free flow of data, as well as preventing the national security risks brought about by the misuse of data [15]-[18].

Based on numerical analysis methods, this study constructs a multiple linear regression model to systematically analyze the influence mechanism of the level of cross-border data flow on the legal regulatory effect of personal information protection. By collecting multi-country panel data and adopting econometric methods such as stepwise regression, endogeneity test, and robustness test, key factors affecting the legal regulatory effect of personal

information protection are identified, which provide scientific basis and policy suggestions for optimizing the cross-border data governance system and enhancing the level of personal information protection.

## II. Regulatory Approach to the Supervision of Cross-Border Movement of Personal Information

The cross-border flow of personal information is the inevitable result of the development of Internet technology and global information exchange. Under the overall security concept, it is not only a civil and commercial relationship between information subjects and information users, but also accompanied by national security risks. China's cybersecurity law and civil code have made basic provisions for the protection of personal information, but there is a lack of legal regulation of the cross-border flow of personal information, especially in cases that may jeopardize cybersecurity and national security.

### *II. A. Definition and significance of the concept*

#### II. A. 1) Definition of concepts

The development of science and technology and changes in business models have transformed the temporary and episodic cross-border transfer of data into widespread and routine migration. In this context, a large number of data trading activities have been generated globally, and different forms and characteristics of the international big data trading market have been formed. In the early discussions on this topic, the concept of “data transfer” was mostly used, but nowadays the concept of “data flow” is mostly used. Nowadays, the concept of “data flow” is used more often. The concept of “data flow” is now more commonly used. Flow is characterized by continuity, and the term “data flow” more accurately expresses the reality of today's data flow, which is fully circulating among countries and regions. With the advent of the big data era, the term “data flow” has been given a new connotation and extension, and has become a new form of network economy. Cross-border flow of personal data is the transfer of personal data to a third country under specific conditions. In the booming development of the Internet and information technology today, people can get a variety of information from many sources, including personal data [19].

There are various forms of cross-border flow of personal information data, which is mainly manifested in the following ways: first, the servers of merchants or websites are located outside the country, and in the process of accessing these merchants or websites, the servers outside the country will keep records of the accesses, and this kind of cross-border flow is a normal, lawful as well as reasonable flow. Secondly, if the merchant or website server is located in China, and some criminals steal the data stored in the domestic server and transmit it to the foreign server, this kind of cross-border flow is illegal and improper flow. Third, a parent company that has set up a subsidiary in China may also generate cross-border flow of data when the overseas parent company transmits the collected personal data of its Chinese customers to the servers of its overseas branches.

This paper considers that the cross-border flow of personal data is divided into two ways, one is the cross-border flow and transmission of personal data, and the second is the cross-border flow of personal data that is not cross-border but can be accessed by subjects in a third country.

#### II. A. 2) Significance

In-depth analysis of the legal regulation of cross-border data flow of personal information, which is of very far-reaching and significant significance for the maintenance of national security, the protection of citizens' personal information security, and the promotion of the development of the country's digital economy [20].

(1) Maintaining national security

Data security is highly related to national security, and data security has been integrated into the national security system. With the emergence and popularization of emerging technologies such as big data and cloud computing, the generation and cross-border flow of data have become more and more simple and rapid, and the current scale of cross-border flow of personal information around the world has been exploding, while cross-border activities of data have posed a significant pressure and risk of leakage for data security supervision due to the characteristics of the process of multilinking, transjurisdictionality, and difficulties in traceability. Therefore, we must strengthen data security supervision, improve relevant laws and regulations, incorporate cross-border activities of personal information into a safe and orderly legal framework, and upgrade relevant technical means to cope with the risks posed by cross-border activities of data.

(2) Protection of Citizens' Personal Information Security

Personal information has its own specificity, and in addition to the property attributes shown as factors of production in the exploitation of data resources and the development of digital trade, it also has distinctive personal attributes. Personal information originates from individuals themselves and their activities, so the safety of personal information is related to the peace of life, reputation, honor and even more so to the property and personal safety of individuals.

(3) Promoting the development of digital economy in the country

At present, the digital economy has become an important and indispensable part of the international economic system, data is known as the “new oil” of the twenty-first century, and countries around the world have passed legislation to compete for data resources, and the flow of data for the country and enterprises to create huge business interests.

### *II. B.Regulatory models and legal regulation*

#### II. B. 1) Regulatory model

With the development and popularization of emerging technologies on the Internet, personal information data has become a major resource for countries and enterprises to improve their competitiveness, and the data also has an important impact on people's lives and government management. Illegal interference in personal information data by any subject may jeopardize the interests of the state, society and citizens of other countries. Based on different policy considerations, countries have respectively adopted different regulatory paths for the cross-border flow of personal information data [21]. The global regulatory models for cross-border data flow of personal information are specified as follows:

(1) Unconditional flow mechanism. Data can flow freely across borders without special requirements.

(2) Conditional flow. Data is prohibited from being transferred abroad unless certain conditions are met by the data receiving country, the data controller or the data processor.

(3) Data localization measures. Data localization measures are measures to restrict the cross-border flow of data, specifically including data local storage and data local processing. Local storage of data means that certain data are prohibited from being transferred across borders unless a copy is stored in the home country. Local processing of data means that the main processing of data must be carried out in the country's data centers, enterprises must establish or rent data centers in the country, or switch to a local data processing service provider.

(4) Prohibition of data transfer. Data must be stored, processed and accessed within the territory of the home country. This type differs from local processing in that data localization does not imply an absolute ban on the free flow of data across borders, whereas a ban on data transfer does not even allow an enterprise to send copies of its data abroad.

#### II. B. 2) Legal regulation

China's personal information data in cross-border data flow regulation faces many practical challenges, we need to constantly review the international development dynamics, combined with their own development needs, to formulate cross-border data flow response programs and strategies, and to strengthen the legal regulation and supervision of data at the source of cross-border data.

(1) Improve the jurisdiction of cross-border flow of personal data and strengthen the construction of internal audit mechanism for cross-border flow security

Through the examination of domestic legislation on cross-border flow of personal data, however, based on the special characteristics of personal data, the risk of cross-border scenarios and other characteristics, personal privacy and autonomy in the process of cross-border flow of data have not yet been able to be effectively protected. The state should set up a diversified categorization mechanism and further clarify the protection standards. In the legislation, it is necessary to further clarify the concepts of “critical infrastructure operators” and “important core data”, and set different data output audit requirements and regulatory standards according to the source and importance of the data in cases involving national security, corporate interests and the rights and interests of personal data. and regulatory standards. Within the framework of national and enterprise cross-border data transmission security management system, implement pilot projects on cross-border data transmission security management, and promote policy synergy in cross-border data legislation, law enforcement and foreign cooperation.

(2) Actively promote the construction of a pilot data free port and create a special zone for cross-border data flow.

Explore the pilot data free port center with “specific region, specific subject, special authorization and special supervision” as the pilot idea, create an experimental zone for domestic and foreign data by carrying out activities such as international data processing and exchanging, and promote the early and pilot implementation of cross-border data flow in the free trade port. Leveraging the new advantages of the international data industry, building a data free port with international influence, and constructing an international digital product exchange platform. Targeting major international trading partners and countries of origin and destination of tourists, it will prioritize the exploration of cross-border data flow in tourism, trade and other fields, make full use of geographic advantages, supplemented by a variety of energy sources, and set up regional cross-border data flow rules and whitelisting mechanisms. Realize the construction of an open, transparent, secure and controllable regulatory system for cross-border data flow in the data free port pilot, and promote the liberalization and facilitation of the digital business

environment around the digital trade barriers constraints and policy obstacles, leading the accelerated development of digital trade.

(3) Increase investment in the supply of “core technologies” and enhance technical guarantee for data security.

Strengthening data security technology is one of the prerequisites for the free flow of data across borders. By strengthening data security technology, illegal data flow across borders can be prevented and restricted. By strengthening data security technology, illegal cross-border data flow can be prevented and restricted, and by adopting appropriate technical measures, timely response to sudden network attacks can be made to prevent data leakage. In response to the various security risks that may arise in cross-border data flows, such as data leakage and misuse, technological innovations should be continuously strengthened in order to provide a reliable environment and management capabilities for cross-border data flows.

## III. Impact model of legal regulation of personal information protection

With the development of the information age, the cross-border flow of personal data is becoming more and more frequent, and data has become an important resource for the competition between countries and gradually evolved into the core of the international strategic game. In order to further analyze the impact of legal regulation on personal data protection on cross-border data flow, this paper conducts numerical analysis based on regression model, aiming to safeguard the puppy of personal data protection in cross-border data flow, and to enhance China's discourse power in global data governance.

### *III. A. Multiple linear regression models*

#### III. A. 1) Basic knowledge

Multiple Linear Regression (MLR) is a widely used analytical model for machine learning, which creates a mathematical model to analyze the value of the dependent variable through the linear relationship between multiple features and the dependent variable. Multiple linear regression models can handle linear relationships between multiple independent variables, which can reduce redundancy between variables and improve the accuracy of model analysis. Meanwhile, the MLR model also has strong interpretability, and the regression coefficients can be used to assess the degree of influence of different independent variables on the dependent variable.The MLR model can also deal with missing data to avoid outliers in data processing affecting the performance of the model. Therefore, this study analyzes the specific impact of the level of cross-border data flow on the effect of legal regulation on personal information protection through multiple linear regression models.

Assuming that there are $n$ independent variables $x_1, x_2, \ldots, x_n$ and a dependent variable $y$, the multiple linear regression model is expressed as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \varepsilon \tag{1}$$

where $\beta_0, \beta_1, \beta_2, \cdots, \beta_n$ are the model coefficients, and $\varepsilon$ is the error term.

The model coefficients reflect the degree of influence of the independent variables on the dependent variable, while the error term reflects factors that cannot be fully explained by the model. The multiple linear regression model can be regarded as an extension of the one-way linear regression model when $\beta_2 = \beta_3 = \cdots = \beta_m = 0$ in the above equation.

In practice, assuming that there are n groups of observations, in which the number of independent variables in each group is m. In this case, the multiple linear regression model can be expressed as follows:

$$\begin{cases} y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_m x_{1m} + \varepsilon_1 \\ y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_m x_{2m} + \varepsilon_2 \\ \vdots \\ y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_m x_{nm} + \varepsilon_n \end{cases} \tag{2}$$

In the practical application of the analysis, the assumptions of the multiple linear regression model include:

(1) Linear relationship. There should be a linear relationship between the level of cross-border data flow and the various factors of the legal regulatory effect of personal information protection.

(2) Independent and identically distributed errors. The errors of the model should be independent of each other and obey the same distribution, i.e. there is no multicollinearity problem.

(3) No multicollinearity. There should be no multicollinearity between independent variables, i.e., there should be no high correlation between independent variables.

(4) Equal error variance. The error variances of the model should be equal.

(5) Normally distributed errors. The errors of the model should follow a normal distribution.

These assumptions must be satisfied when using multiple linear regression models for numerical analysis, otherwise they may lead to inaccurate regression results.

**III. A. 2) Parameter estimation**

When setting up a multiple regression model for sample data $(x_i, y_i)(i = 1,2,3\ldots m)$ and wanting to find the minimum value $a_0, a_1, a_2, a_2 \ldots a_n$ of $b_0, b_1 \ldots b_n$ the least squares method should be used, i.e., the sum of the squares of the errors is minimized. That is:

$$\sum_{i=1}^{m} \varepsilon^2 = \sum_{i=1}^{m} (y_i - f(x_i; b_1 b_2 \ldots\ldots b_m))^2 \tag{3}$$

Min. When $b_0, b_1 \ldots b_n = a_0, a_1, a_2, a_2 \ldots a_n$ ], the above formula reaches its minimum, and it is said that $a_0, a_1, a_2, a_2 \ldots a_n$ is the point estimate of $b_0, b_1 \ldots b_n$ . Then the regression estimate of the variable y can be written as:

$$\hat{y}_n = a_0 + a_1 x_{n1} + a_2 x_{n2} + \ldots a_m x_{nm} \tag{4}$$

At this point, remember:

$$\delta = y_n - \hat{y}_n \tag{5}$$

is the residual. $\sum_{i=1}^{n} \delta^2$ is the residual sum of squares, denoted RSS. Notation:

$$S = \frac{RSS}{(n-k)} \tag{6}$$

is called the residual variance, while S is called the residual standard deviation. Regression analysis is a mathematical tool for studying correlations through conditional expectations, and therefore establishes a relationship between the actual, predicted and mean values. Namely:

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 - \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \tag{7}$$

That is, Total Sum of Squares (TSS) - Regression Sum of Squares (RSS) = Residual Sum of Squares (ESS). Then:

$$\sum_{i=1}^{n} (y_i - \bar{y})^2 - \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \tag{8}$$

The goal should be to achieve the best fit by making the residual sum of squares and residual standard deviation reach the target minimum.

To determine whether the effect of each independent variable on the dependent variable is significant or not, it is necessary to perform a t-test for each regression coefficient B, i.e.:

$$t = \frac{\hat{B}}{S} \sim t(n-k-1) \tag{9}$$

where S is the standard deviation of the sampling distribution of B.

***III. B. Variable selection and modeling***

**III. B. 1) Variable selection**

(1) The explanatory variable is the effect of legal regulation on personal information protection (GRXX). According to the statistical method of the United Nations Conference on Trade and Development on the effect of legal regulation on personal information protection, the types of personal information data that can be delivered across borders are selected, and the data on the effect of legal regulation on personal information protection are calculated from this.

(2) The core explanatory variable is the level of cross-border data flow (KJSJ). For the measurement of the level of cross-border data flow, this paper uses the DSTRI proposed by the OECD as a proxy variable for the liberalization of the level of cross-border data flow in each country. The specific value of the index is between 0 and 1, with higher values indicating more liberalized cross-border data flow levels.

(3) Control Variables. For the influencing factors of the legal regulatory effect of personal information protection, this paper mainly analyzes the level of economic development (GDP), geographic distance (DIS), population size (POP), the level of Internet infrastructure development (INT), and the level of foreign direct investment (FDI).

Because the OECD DSTRI database is for the period 2016-2024, and noting the accessibility of the legal regulatory effects of personal information protection, this paper selects panel data for 12 exporting countries and 30 importing countries for the sample time period 2016-2024 for the regression. The source of the other control variables is the CEPII database.

**III. B. 2) Modeling**

In order to study the impact of the level of cross-border data flow on the effect of legal regulation on personal information protection, this paper takes the multiple linear regression model as the basis, incorporates the relevant variables that may affect the effect of legal regulation on personal information protection, and sets the benchmark regression model as follows:

$$GRXX_{ijt} = \alpha_0 + \alpha_1 KJSJ_{ijt} + \alpha_2 Controls_{ijt} + \lambda_j + \lambda_t + \varepsilon_{ijt} \tag{10}$$

where $i$ denotes China, $j$ denotes countries and regions that have signed trade agreements with China, $t$ is time, and $\alpha$ are all parameters to be estimated. $GRXX_{ijt}$ is the explanatory variable, $KJSJ_{ijt}$ is the core explanatory variable, and $Controls_{ijt}$ denotes the control variables, including the level of economic development, geographic distance, population size, the level of Internet development, and the level of foreign direct investment. $\lambda_j$ denotes region fixed effects, $\lambda_t$ denotes year fixed effects, and $\varepsilon_{ijt}$ denotes random error terms.

## IV. Factors affecting the legal regulation of personal information protection

In recent years, with the widespread application of Internet technology and big data technology and the rapid development of digital economy and trade, the importance of data has become more and more prominent, and international economic exchanges have increasingly relied on data as a medium to promote the development of international economy and trade. However, against the background of increasingly frequent cross-border flow of data, data leakage, data misuse, data poisoning and other incidents that jeopardize data security are frequent, sounding the alarm for legislation in the field of data security. Among them, personal data, as the most important part of data, is closely related to the legal rights of individuals, while the improper utilization of personal data also poses a serious threat to national security, so it is necessary to improve the legal regulation of cross-border flow of personal data.

### *IV. A. Correlation test*

The correlation test is a statistical test of whether there is a correlation between variables and the strength and direction of the correlation. The value of correlation coefficient is between -1 and 1, -1 and 1 indicate perfect negative correlation and perfect positive correlation, respectively, and synchronous change between two variables, while 0 indicates that there is no correlation between variables. The correlation test can not only reflect the correlation between the explanatory variables and the explained variables but also test whether there is a problem of covariance between the independent variables. Table 1 shows the results of the correlation test, in which *,**,*** represent the significance at 10%, 5% and 1% levels respectively. As can be seen from the table, there is a significant positive correlation between the core explanatory variable cross-border data flow level (KJSJ) and each of the control variables with the explanatory variable personal information protection legal regulation effect (GRXX). The correlation coefficients between the control variables INT and DIS, FPOP and DI are large.

In order to avoid the distortion or difficulty in estimating the model due to the existence of precise correlation or high correlation between explanatory variables in the model, it is necessary to combine with the Variance Inflation Factor (VIF) to test whether there is multicollinearity between explanatory variables. Heteroskedasticity means that the random error terms have different variances, which is relative to homoskedasticity, the so-called homoskedasticity means that the random error terms in the overall regression function satisfy the homoskedasticity, which is to ensure that the regression parameter estimates have good statistical properties. In order to ensure that the regression parameter estimates have good statistical properties, it is necessary to carry out the heteroskedasticity test on the panel data. The original hypothesis of the test is that the random errors of the

regression equation satisfy homoskedasticity. The results of the test of multicollinearity and heteroskedasticity test are shown in Table 2.

Table 1: Correlation test results

| | GRXX | KJSJ | GDP | DIS | POP | INT | FDI |
|---|---|---|---|---|---|---|---|
| GRXX | 1.000 | - | - | - | - | - | - |
| KJSJ | 0.412*** | 1.000 | - | - | - | - | - |
| GDP | 0.528*** | 0.262*** | 1.000 | - | - | - | - |
| DIS | 0.636*** | 0.428*** | 0.419*** | 1.000 | - | - | - |
| POP | 0.443*** | 0.174*** | 0.174* | 0.227*** | 1.000 | - | - |
| INT | 0.557*** | 0.286*** | 0.428*** | 0.514*** | 0.134 | 1.000 | - |
| FDI | 0.374*** | 0.105*** | 0.311*** | 0.043 | 0.728*** | 0.185* | 1.000 |

The test results of multicollinearity show that the VIF values of the explanatory variables, core explanatory variables and control variables are all less than the critical value of 5, and the mean value of the variance inflation factor is also less than the critical value of 5. Therefore, there is no problem of covariance between the explanatory variables selected in this paper, and regression analyses can be performed. In addition, the heteroskedasticity results show that there is heteroskedasticity in the data, indicating that robust standard errors are needed later to eliminate heteroskedasticity.

Table 2: Multiple colinear and isovariance test

| Variable | VIF | 1/VIF | Isovariance test |
|---|---|---|---|
| GRXX | 4.671 | 0.214 | chi2(1) =38.217 |
| KJSJ | 4.528 | 0.221 | |
| GDP | 4.316 | 0.232 | |
| DIS | 4.105 | 0.244 | |
| POP | 3.792 | 0.264 | Prob > chi2= 0.001 |
| INT | 3.267 | 0.306 | |
| FDI | 2.849 | 0.351 | |
| Means | 3.933 | 0.254 | |

### *IV. B. Benchmark regression analysis*

In order to analyze the influence of the level of cross-border data flow on the effect of legal regulation on personal information protection, based on the multiple linear regression model constructed in the previous section, this paper conducts a stepwise regression analysis with data from multiple countries. The results of the stepwise regression analysis are shown in Table 3, where the values in parentheses are the standard deviation, and *,**,*** represent the significance at the 10%, 5%, and 1% levels, respectively, and the same as below.

In order to observe whether the level of cross-border data flow (KJSJ) has a driving effect on the effect of legal regulation on personal information protection (GRXX), in the first column of the regression, this paper only includes the level of cross-border data flow (KJSJ) as an explanatory variable. The regression results show that the coefficient of the level of cross-border data flow is significantly positive at the 1% level, and its regression coefficient reaches 0.792, indicating that the increase of the level of cross-border data flow can promote the national legal regulation effect of personal information protection, i.e., for every increase of 1 percentage point in the level of cross-border data flow, the effect of the legal regulation of personal information protection will be enhanced by 0.792 percentage points. In addition, the regression results of model (2) to (5) are obtained after gradually adding the control variables selected in the previous section. The results show that after gradually adding control variables, the coefficient of cross-border data flow level (KJSJ) shows a gradual decline compared with the regression results of model (1), and the regression coefficients of the influence of each control variable on the effect of legal regulation of personal information protection are significantly positive. It indicates that the level of cross-border data flow of a country affects the legal regulation effect of personal information protection by influencing the country's level of economic development (GDP), geographic distance (DIS), population size (POP), the level of development of Internet infrastructure (INT) and the level of foreign direct investment (FDI). In addition, the $R^2$ coefficient of determination of the baseline regression result of the model is 0.603 before adding control variables, while the $R^2$ coefficient of determination of the model increases to 0.914 after adding each control variable, which gradually

improves the explanatory power of the model. It shows that the effect of legal regulation on personal information protection is affected by many different factors, which provides a direction for further optimizing the effect of legal regulation on personal information protection in cross-border data flow.

Table 3: Benchmark regression analysis results

| Variable | Model (1) | Model (2) | Model (3) | Model (4) | Model (5) |
|---|---|---|---|---|---|
| KJSJ | 0.792*** (0.084) | 0.473*** (0.079) | 0.428*** (0.075) | 0.329*** (0.071) | 0.307*** (0.085) |
| GDP | - | 0.248*** (0.051) | 0.215*** (0.063) | 0.232*** (0.054) | 0.172** (0.055) |
| DIS | - | 0.233** (0.059) | 0.271*** (0.054) | 0.103* (0.045) | 0.081*** (0.041) |
| POP | - | - | - | 0.272*** (0.063) | 0.195*** (0.059) |
| INT | - | - | - | 0.057*** (0.009) | 0.122*** (0.031) |
| FDI | - | - | - | - | 0.201** (0.063) |
| (Con_) | 8.971*** (0.927) | 8.562*** (0.638) | 8.041*** (0.749) | 8.573*** (0.735) | 7.195*** (0.814) |
| Time | YES | YES | YES | YES | YES |
| Region | YES | YES | YES | YES | YES |
| $R^2$ | 0.603 | 0.806 | 0.838 | 0.872 | 0.914 |

### *IV. C. Endogeneity test*

In the benchmark regression of this paper, after adding other control variables, controlling for regional fixed effects and time fixed effects, the regression of the model can, to a certain extent, reduce the endogeneity problem in the model and improve the accuracy and reliability of the estimated coefficients of the model. However, there may still be potential endogeneity problems in the model, and the consideration of endogeneity in this paper lies in the possible bidirectional causality between the core explanatory variables and the explanatory variables of the model, i.e., on the one hand, the increase in the level of cross-border data flow can promote the increase in the effect of the legal regulation of personal information protection, and on the other hand, the increase in the effect of the legal regulation of personal information protection may also force the increase in the level of cross-border data flow. The relationship between the possible bidirectional causality between the explanatory variables and the core explanatory variables can make the estimation of the model coefficients biased, resulting in unreliable estimated coefficients of the model, so this paper provides further discussion on the endogeneity issue.

The treatment of endogeneity in this paper is mainly carried out in the following two ways:

(1) Based on the instrumental variable approach, this paper draws on existing studies to use the one-period lagged term of the cross-border data flow level indicator as an instrumental variable for regression. In addition, the instrumental variables of the model passed the unidentifiable test, weak instrumental variable test and over-identification test, which verified the rationality of the instrumental variables.

(2) In this paper, with reference to the practice of related studies, generally speaking, if there is no obvious endogeneity problem in the model, the change of the future legal regulation effect of personal information protection should be uncorrelated with the change that occurs in the current period, so after adding the front term of cross-border data flow level to the model, the front term does not have a significant effect. The regression analysis of the model is carried out after adding the one-period front term of the digital trade terms level indicator.

Table 4 shows the test results of the endogeneity analysis. The results in Column (1) show that after using the lagged one-period term of the cross-border data flow level as an instrumental variable, the estimated coefficient of the model's cross-border data flow level (KJSJ) is still significantly positive at the 1% level, i.e., every 1% increase in the level of cross-border data flow can make the effect of the personal information protection law regulation rise by 7.9%, which is basically the same as the baseline regression results. The results in column (2) show that after adding the one-period prior term of cross-border data flow level, the estimated coefficient of the prior term is not significant, and the estimated coefficient of the current term is significantly positive at 1% level, which verifies the robustness of the model regression results.

Table 4: Endogeneity analysis test results

| - | Model (1) | Model (2) |
|---|---|---|
| $L_{n-1}$ (KJSJ) | 0.079***(0.007) | 0.048**(0.005) |
| $L_{n+1}$ (KJSJ) | - | 0.006(0.003) |
| Kleibergen-Paap rk LM statistic | 581.327(0.000) | - |
| Kleibergen-Paap rk Wald F statistic | 1672.839(20.131) | - |
| Hansen J statistic | 0.185(0.627) | - |
| Control variable | YES | YES |
| Region effect | YES | YES |
| Time effect | YES | YES |
| $R^2$ | 0.479 | 0.507 |

### *IV. D. Robustness Tests*

The robustness tests in this paper are divided into two categories, namely, the robustness tests of expanded sample size and indicator measures. The robustness test of expanding sample size is to add the data of personal information protection awareness (XXBH) to the original sample data, and the robustness test of indicator metrics cites the Digital Services Trade Restriction Index (DSTRI) published by OECD to put in the robustness of regression. Table 5 shows the results of the robustness test.

The regression coefficients of the core explanatory variables in Column (1) of Table 5 remain positive, confirming that the level of cross-border data flow enhances the effect of legal regulation on personal information protection, and that the higher the level of a country's economic development and the deeper its openness to the outside world, the more the export sector of data-intensive trade in services receives its positive impact. While the results of column (2) after adding all control variables are quite different from the original model results, and the impact coefficient of personal information protection awareness (XXBH) is 0.176, which is significantly positive at 1% level. The reason for this may be because although personal information protection awareness relies more on data, it involves less national privacy, company secrets, etc., and the state's restrictive measures in these areas are much less in comparison to other digital service sectors. Second, with the rise of the digital wave, the demand for new lifestyles and digital leisure and entertainment continues to grow. The level of cross-border data flows continues to climb, thus contributing more significantly to the effect of legal regulation on personal information protection.The OECD publishes the Digital Services Trade Restrictions Index (DSTRI), and when a country scores higher, it indicates that the country has more measures related to the level of cross-border data flows, and thus contributes more to the effect of legal regulation on personal information protection. Therefore, it is necessary to add the new variable DSTRI to test the smoothness of the baseline regression model. The results in Column (3) of Table 5 show that after adding the DSTRI variable, the sign of the core explanatory variable, cross-border data flow level, remains positive and statistically significant, and the control variables are also as expected. That is, the sign and significance of the overall parameters remain consistent with the previous paper, which further proves the reliability of the findings of this paper.

Table 5: Robustness test results

| Variable | Model (1) | Model (2) | Model (3) |
|---|---|---|---|
| KJSJ | 0.581***(0.082) | 0.247***(0.046) | 0.209***(0.075) |
| GDP | - | 0.101**(0.037) | 0.075**(0.041) |
| DIS | - | 0.075***(0.052) | 0.069***(0.038) |
| POP | - | 0.083***(0.061) | 0.058***(0.027) |
| INT | - | 0.128***(0.049) | 0.104***(0.052) |
| FDI | - | 0.147**(0.038) | 0.136**(0.048) |
| XXBH | - | 0.176***(0.051) | - |
| DSTRI | - | - | 0.181***(0.063) |
| (Con_) | 6.785***(0.343) | 8.542***(0.175) | 10.641***(0.471) |
| Time | YES | YES | YES |
| Region | YES | YES | YES |
| $R^2$ | 0.547 | 0.316 | 0.446 |

## V. Conclusion

This study reveals the significant impact of the level of cross-border data flows on the legal regulatory effect of personal information protection based on the empirical analysis of multiple linear regression models. The core findings indicate that the increase in the level of liberalization of cross-border data flows can significantly contribute to the enhancement of the legal regulatory effect of personal information protection, and there is a stable positive relationship between the two. Specifically, the model coefficient of determination increases from 0.603 in the base regression to 0.914 in the full model, indicating that the selected variables can better explain the changes in the legal regulation effect of personal information protection. The endogeneity test further confirms the reliability of this finding, with the estimation results under the instrumental variables approach showing that every 1% increase in the level of cross-border data flow can lead to a 7.9% increase in the legal regulatory effect of personal information protection. In the robustness test, the impact coefficient of the personal information protection awareness variable reaches 0.176 and is statistically significant, verifying the robustness of the findings. These findings suggest that when countries formulate cross-border data governance policies, they should take into full consideration the combined effects of multiple factors, such as the level of economic development, infrastructure construction, and population size, and build a governance system that protects the security of personal information and facilitates the flow of data elements by improving the legal regulatory framework, strengthening international cooperation, and upgrading the capacity of technological safeguards.

## Funding

This work was supported by Anhui Provincial Department of Education Humanities Key Project: "Empirical Approach to Criminal Compliance Risk Patterns and Decriminalization of Small and Medium sized Enterprises" (2024AH053234).

## References

[1] Meltzer, J. P. (2015). The I nternet, Cross-Border Data Flows and International Trade. Asia & the Pacific Policy Studies, 2(1), 90-102.
[2] Tuthill, L. L. (2016). Cross-border data flows: What role for trade rules?. In Research handbook on trade in services (pp. 357-382). Edward Elgar Publishing.
[3] Voss, W. G. (2019). Cross-border data flows, the GDPR, and data governance. Wash. Int'l LJ, 29, 485.
[4] Aaronson, S. A. (2019). Data is different, and that's why the world needs a new approach to governing cross-border data flows. Digital Policy, Regulation and Governance, 21(5), 441-460.
[5] Chin, Y. C., & Zhao, J. (2022). Governing cross-border data flows: International trade agreements and their limits. Laws, 11(4), 63.
[6] LeSieur, F. (2012). Regulating cross-border data flows and privacy in the networked digital environment and global knowledge economy. Int'l Data Priv. L., 2, 93.
[7] Sun, L. (2023). Overview of regulations on cross border data flow. Academic Journal of Science and Technology, 8(1), 171-176.
[8] Ahmed, U., & Chander, A. (2015). Information Goes Global: Protecting Privacy, Security, and the New Economy in a World of Cross-border Data Flows. In E15Initiative. Geneva: International Centre for Trade and Sustainable Development (ICTSD) and World Economic Forum.
[9] Tehrani, P. M., Sabaruddin, J. S. B. H., & Ramanathan, D. A. (2018). Cross border data transfer: Complexity of adequate protection and its exceptions. Computer law & security review, 34(3), 582-594.
[10] Ziyi, X. (2022). International Law Protection of Cross-Border Transmission of Personal Information Based on Cloud Computing and Big Data. Mobile Information Systems, 2022(1), 9672693.
[11] Chen, M. (2024). Developing China's Approaches to Regulate Cross-border Data Transfer: Relaxation and Integration. Computer Law & Security Review, 54, 105997.
[12] Ferracane, M. F. (2017). Restrictions on cross-border data flows (No. 01/2017). ECIPE working paper.
[13] Singh, S. (2024). Regulation of Cross-Border Data Flow and Its Privacy in the Digital Era. NUJS J. Regul. Stud., 9, 38.
[14] Rong, K., Ling, Y., Yang, T., & Huang, C. (2025). Cross-border data transfer: patterns and discrepancies. Journal of International Business Policy, 1-23.
[15] Dai, L. (2019). A Survey of Cross-Border Data Transfer Regulations through the Lens of the International Trade Law Regime. NYUJ Int'l L. & Pol., 52, 955.
[16] Alhababi, H. H. (2024). Cross-Border Data Transfer between the gcc Data Protection Laws and the gdpr. Global Journal of Comparative Law, 13(2), 178-200.
[17] Dayday, C. M. G. T. (2023). Cross-border data flows and data regulation under international trade law. Phil. LJ, 96, 33.
[18] Tahir, S., & Tahir, W. (2024). Legal Challenges in Cross-Border Data Transfers: Balancing Security and Privacy in a Globalized World. Mayo Communication Journal, 1(1), 1-11.
[19] Guanxiang Wang. (2023). The Necessity of Extraterritorial Application of the Personal Information Protection Law and Suggestions for Its Improvement. International Journal of Frontiers in Sociology,5(16).
[20] Ziyi Xu. (2022). International Law Protection of Cross-Border Transmission of Personal Information Based on Cloud Computing and Big Data. Mobile Information Systems,2022.
[21] Jung Aeryung. (2019). Review of requirements for transborder flow of personal information. Public Law Journal,20(2),209-244.