

Developing a Model for Big Data–Based Tax Collection and Management in an Integrated Tax System and Its Impact on Corporate Investment Inefficiency

Mohsen Keymanesh ¹, Khadijeh Ebrahimi Kahrizsangi ^{2,*} and Arezoo Aghaei Chadegan ³



¹ Department of Accounting, Na.C., Islamic Azad University, Najafabad, Iran; 

² Department of Accounting, Na.C., Islamic Azad University, Najafabad, Iran; 

³ Department of Accounting, Na.C., Islamic Azad University, Najafabad, Iran; 

* Correspondence: Ebrahimi641@iaun.ac.ir

Citation: Keymanesh, M., Ebrahimi Kahrizsangi, K., & Aghaei Chadegan, A. (2027). Developing a Model for Big Data–Based Tax Collection and Management in an Integrated Tax System and Its Impact on Corporate Investment Inefficiency. *Business, Marketing, and Finance Open*, 4(3), 1–16.

Received: 02 March 2026

Revised: 13 July 2026

Accepted: 20 July 2026

Initial Publication: 21 August 2026

Final Publication: 01 May 2027



Copyright: © 2027 by the authors. Published under the terms and conditions of Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

Abstract: A big data–based tax collection and management system relies on large-scale datasets shared among tax authorities, companies, and government agencies. By drawing on new-generation information technologies, such as big data analytics and cloud computing, this system improves data analysis methods through the comparison, analysis, cross-checking, and examination of tax-related information, thereby enabling the effective prediction of tax risks and the provision of multidimensional, in-depth insights. The present study aimed to develop a model for big data–based tax collection and management within an integrated tax system and to examine its impact on corporate investment inefficiency. The study adopted a qualitative approach and used semi-structured interviews with experts in taxation and big data. Data were analyzed using MAXQDA software based on the grounded theory approach proposed by Strauss and Corbin. Following open, axial, and selective coding, the findings were organized into principal categories and themes. The results indicated that data quality and standardization are among the most important prerequisites for establishing a data-driven taxation system. The absence of uniform standards and poor data quality reduce analytical accuracy and disrupt financial decision-making. Institutional integration and the development of effective mechanisms for interorganizational information exchange were also identified as key factors in reducing tax evasion and enhancing transparency. Information transparency and data security were recognized as essential components for increasing economic actors' trust and improving the efficiency of the taxation system. Furthermore, strengthening corporate governance, improving internal control systems, and reforming legal frameworks were identified as supporting requirements for establishing a big data–based taxation system.

Keywords: Big data taxation, integrated taxation system, investment inefficiency.

1. Introduction

Taxation is one of the principal mechanisms through which governments mobilize financial resources, finance public services, regulate economic activity, redistribute income, and support social and developmental objectives. The effectiveness of a tax system therefore depends not only on tax rates and statutory provisions but also on the institutional capacity of tax administrations to identify taxpayers, assess taxable activities, collect accurate information, detect noncompliance, and enforce tax obligations fairly. In contemporary economies, these functions have become increasingly complex because business transactions are conducted through interconnected digital platforms, financial institutions, electronic marketplaces, and cross-organizational networks. Traditional tax

collection approaches, which are largely based on periodic declarations, fragmented databases, manual audits, and retrospective controls, are often unable to process the scale, speed, and variety of modern economic data. Consequently, governments increasingly require integrated, transparent, and technologically advanced tax administration systems capable of converting large volumes of heterogeneous data into reliable tax intelligence. From a broader governance perspective, effective tax resource collection and allocation must also be grounded in transparency, accountability, fairness, and the protection of public interests because the legitimacy of fiscal institutions is closely associated with the equitable mobilization and use of financial resources [1, 2].

The rapid expansion of digital technologies has created new opportunities for transforming tax administration from a document-based and reactive system into a data-driven and preventive system. Big data technologies enable tax authorities to collect, integrate, store, and analyze information generated by banking transactions, customs records, insurance databases, electronic invoices, corporate reports, property registries, payroll systems, and online commercial activities. Through such integration, tax administrations can establish more comprehensive profiles of taxpayers, identify inconsistencies among reported and actual economic activities, and allocate audit resources according to measurable risk. Evidence from the Serbian tax administration indicates that big data analytics can improve tax collection efficiency by enabling more systematic taxpayer segmentation, risk assessment, and identification of irregularities that may not be visible through conventional audit procedures [3]. In this context, big data are not merely a larger volume of information; rather, they represent an analytical infrastructure through which dispersed administrative records can be converted into actionable knowledge for tax assessment, compliance supervision, and revenue management.

The development of an integrated taxation system requires more than the acquisition of advanced software. It depends on the quality, compatibility, completeness, timeliness, and accessibility of data collected from multiple public and private institutions. When databases are designed independently, use different definitions and coding structures, or contain incomplete and inaccurate records, their integration may produce misleading analytical results. The problem is particularly significant in tax administration because errors in identifying individuals, companies, transactions, and ownership relationships can influence audit selection, tax assessments, penalties, and investment decisions. Studies of electronic tax collection have consequently emphasized that technological infrastructure, data integration, organizational coordination, human resources, legal requirements, and taxpayer-related factors must be considered together when designing comprehensive electronic tax systems [4, 5]. A reliable big data-based tax system must therefore include uniform data standards, interoperable databases, shared identification protocols, continuous data-quality controls, and clear rules governing the responsibilities of institutions that produce or exchange fiscal information.

Institutional fragmentation is one of the most substantial barriers to the establishment of such a system. Tax authorities frequently depend on information held by banks, customs agencies, social insurance organizations, municipalities, companies, stock exchanges, and other governmental bodies. However, administrative resistance, incompatible information systems, concerns regarding confidentiality, and ambiguous legal mandates may prevent the timely exchange of these data. Research on the organizational challenges of tax collection in Iran demonstrates that tax administration involves multiple structural and managerial complexities, including coordination problems, procedural weaknesses, administrative obstacles, and difficulties in managing the interactions among taxpayers, tax officials, and governmental institutions [6]. These challenges suggest that an integrated tax system cannot be established solely through technological modernization. It also requires institutional redesign, clearly defined data-

governance arrangements, interorganizational agreements, managerial commitment, and mechanisms for resolving conflicts over data ownership, access, and accountability.

The transition toward electronic and intelligent tax administration is particularly important for controlling tax evasion. Tax evasion arises when taxpayers intentionally conceal income, underreport transactions, manipulate expenses, conduct fictitious activities, or otherwise avoid legally required tax payments. Such conduct reduces public revenue, creates unfair competition, weakens trust in government, and transfers the tax burden to compliant taxpayers. Behavioral and institutional factors may jointly influence tax evasion because noncompliance can be affected by perceived fairness, enforcement capacity, personal motivations, expected penalties, and opportunities for concealment. Evidence concerning psychological egoism and tax evasion shows that taxpayer behavior can materially influence revenue collection performance and that improving collection outcomes requires attention to both administrative controls and the motivational foundations of noncompliance [7]. Accordingly, a comprehensive tax management model should combine technological detection capabilities with legal enforcement, taxpayer education, ethical governance, and measures that strengthen the perceived legitimacy of the tax system.

Electronic tax collection can reduce opportunities for concealment by replacing fragmented paper-based procedures with traceable digital records. It can also facilitate automated cross-checking between declarations, financial transactions, and third-party reports. A comprehensive electronic tax collection model based on interpretive structural modeling has identified the importance of interconnected technological, legal, organizational, and supervisory components in reducing tax evasion [8]. Similarly, research on smart governance has emphasized that intelligent tax collection contributes to the control of tax evasion by improving information transparency, accelerating access to taxpayer records, and strengthening the ability of tax authorities to detect inconsistencies [9]. These findings indicate that the effectiveness of digital taxation depends on the extent to which electronic systems are integrated into the broader governance structure rather than being implemented as isolated administrative tools.

Artificial intelligence further expands the analytical capacity of big data-based tax systems. Machine-learning algorithms can process complex datasets, recognize hidden relationships, classify taxpayers according to risk, and identify unusual transaction patterns that may indicate fraud or underreporting. Recent research has shown that artificial intelligence techniques can optimize tax compliance monitoring, fraud detection, and revenue collection by supporting predictive analysis and more precise targeting of enforcement activities [10]. Artificial intelligence systems based on self-organizing maps have also been proposed for value-added tax collection, demonstrating how unsupervised learning can classify complex taxpayer patterns and identify potentially anomalous cases [11]. These applications can reduce dependence on manual audits, increase the speed of detection, and direct limited administrative resources toward taxpayers and transactions with the highest estimated risk. Nevertheless, algorithmic models must be supported by accurate data, transparent decision rules, skilled personnel, and safeguards against biased or erroneous classifications.

Blockchain technology has also been discussed as a complementary instrument for improving the reliability of tax information. By creating tamper-resistant records and facilitating the traceability of transactions, blockchain can strengthen the integrity of data exchanged among taxpayers, financial institutions, and tax authorities. The combined application of blockchain and artificial intelligence has been presented as a means of enhancing tax collection by improving transaction verification, reducing information manipulation, and enabling more intelligent compliance monitoring [12]. However, the practical value of such technologies depends on their alignment with

existing legal systems, organizational capacities, and national data infrastructures. Advanced technologies cannot compensate for weak data governance, inconsistent reporting requirements, or the absence of institutional trust. Therefore, the design of a big data tax model must prioritize technological appropriateness and organizational feasibility rather than technological sophistication alone.

Tax audits, compliance supervision, and enforcement mechanisms remain indispensable even in highly digitalized systems. Electronic information can identify risk indicators, but tax administrations must still possess the authority and professional capacity to investigate anomalies, verify evidence, and impose appropriate corrective measures. Empirical evidence indicates that tax audits, tax collection activities, and the supervision of taxpayer compliance can contribute to increased tax revenues when these mechanisms are implemented coherently [13]. At the same time, enforcement should be consistent with the principles of good governance. Tax sanctions must be proportionate, predictable, legally grounded, and administered without discrimination because arbitrary or excessively punitive enforcement may weaken taxpayer trust and voluntary compliance [2]. The integration of big data analytics with fair enforcement can therefore improve both the effectiveness and legitimacy of tax administration.

The design of tax collection systems must also account for taxpayer satisfaction and acceptance. Digital platforms may reduce administrative costs, shorten processing times, improve access to services, and increase procedural convenience. Research on electronic tax collection has shown that the quality and functionality of electronic systems can influence taxpayers' satisfaction with tax services [14]. Nevertheless, digitalization can also create new forms of exclusion when taxpayers differ in access to technology, digital literacy, connectivity, and the capacity to comply with electronic procedures. Although the digital divide has been widely discussed in educational contexts, its underlying implications concerning unequal access to technological infrastructure, skills, and digital resources are equally relevant to public taxation systems [15]. A technologically advanced tax system that does not provide accessible interfaces, user support, alternative service channels, and adequate training may increase compliance difficulties for small businesses and less digitally capable taxpayers.

Fairness is another central requirement of tax collection. Taxpayers are more likely to accept fiscal obligations when assessment methods are transparent, similarly situated taxpayers are treated consistently, and the tax burden is aligned with actual economic capacity. A model of fair tax collection should therefore integrate legal justice, procedural transparency, accurate assessment, organizational accountability, and protection of taxpayer rights [16]. Tax incentives must also be carefully designed because they may encourage compliance and productive activity but may simultaneously create opportunities for unequal treatment, avoidance, or inefficient resource allocation. Research conducted within the framework of comprehensive tax planning has indicated that tax incentives influence tax collection outcomes and should be evaluated in relation to institutional conditions, taxpayer behavior, and administrative capacity [17]. Big data analytics can support fairer tax administration by enabling authorities to distinguish legitimate economic behavior from opportunistic reporting and to assess the actual distributional effects of exemptions, deductions, and incentives.

The socioeconomic importance of effective tax collection extends beyond government revenue. Tax resources finance infrastructure, education, health care, social protection, and regional development, and weaknesses in collection may constrain governments' ability to reduce inequality and provide essential services. Evidence regarding value-added tax revenue transfers indicates that tax revenues can contribute to regional social development when they are effectively mobilized and allocated [18]. Thus, the performance of tax administration has direct implications for social welfare and public trust. An integrated tax system that reduces evasion and

expands the tax base may allow governments to generate revenue without imposing disproportionate additional burdens on compliant taxpayers. It may also improve the transparency of the relationship between economic activity, tax contributions, and public expenditure.

The architecture of a taxation information system also influences its usability, reliability, and institutional memorability. Although research on space syntax and imageability has primarily examined how structural configurations affect the recognition and recollection of physical spaces, its broader insight—that coherent organization improves comprehensibility and navigation—can be conceptually extended to complex digital systems [19]. A fragmented tax information environment with disconnected databases and inconsistent interfaces may be difficult for both officials and taxpayers to understand and use. By contrast, an integrated information architecture can create clearer pathways for data entry, verification, analysis, and decision-making. This analogy highlights the importance of designing tax systems as coherent organizational environments in which users can identify functions, trace information flows, and understand the consequences of administrative actions.

Beyond tax collection efficiency, the integration of big data into taxation may affect corporate financial decisions, particularly investment efficiency. Investment inefficiency arises when companies invest more or less than the level justified by their economic opportunities. Overinvestment may occur when managers use free cash flow for unproductive projects, related-party transactions, or activities that serve private interests, whereas underinvestment may occur when information asymmetry, financing constraints, or uncertainty prevent firms from undertaking profitable projects. Weak tax transparency can intensify these problems because inaccurate reporting and concealed transactions reduce the ability of investors, regulators, creditors, and boards of directors to assess companies' actual financial conditions. A big data-based taxation system can reduce information asymmetry by cross-checking corporate reports against independent administrative and transactional sources. This improved transparency may restrict fictitious activities, reduce earnings manipulation, strengthen internal controls, and encourage companies to allocate resources toward economically justifiable investments.

The relationship between tax administration and investment efficiency is also mediated by corporate governance. When managers expect that financial transactions, ownership connections, and tax declarations can be systematically analyzed, the perceived probability of detecting opportunistic behavior increases. This may discourage the diversion of resources, fictitious expenditures, and investments designed primarily to conceal income or obtain unjustified tax benefits. At the same time, reliable tax information can improve investors' assessments of corporate risk and performance. The development of electronic tax collection models has already emphasized the importance of combining information infrastructure, organizational controls, legal arrangements, and stakeholder participation [5, 8]. However, the consequences of these systems for corporate investment behavior have received less systematic attention than their effects on tax revenue and compliance.

Existing research has provided valuable evidence regarding big data analytics in tax administration, artificial intelligence for fraud detection, electronic tax collection, taxpayer compliance, tax audits, fair taxation, tax incentives, and organizational challenges. Nevertheless, these topics have often been examined separately. Studies focusing on technological tools may insufficiently address legal and institutional barriers, whereas studies of organizational or behavioral factors may not explain how large-scale data integration can transform monitoring and decision-making. Moreover, the practical success of a comprehensive system depends on the interaction among data quality, technological infrastructure, institutional integration, information security, corporate governance, taxpayer behavior, and regulatory reform. A contextual model is therefore needed to explain the causal conditions,

contextual conditions, intervening factors, strategies, and consequences associated with the establishment of an integrated big data-based tax collection and management system.

Accordingly, the aim of the present study was to develop a model for big data-based tax collection and management within an integrated taxation system and to explain its impact on corporate investment inefficiency.

2. Methodology

This qualitative study was based on semi-structured interviews with experts and specialists in the field of taxation. Data were collected through interviews with 10 managers, tax experts, and faculty members and were reviewed and reassessed over several rounds using the Delphi technique to achieve consensus regarding the factors affecting big data management and the reduction of tax evasion. All interviews were audio-recorded and documented through note-taking and were subsequently imported into MAXQDA software. Open, axial, and selective coding procedures were then performed to extract concepts, categories, and relationships among the variables. MAXQDA facilitated the systematic analysis of qualitative data, the identification of themes, and the development of conceptual diagrams for designing the research model. It also assisted the researcher in organizing and validating the qualitative findings.

Table 1. Statistical Characteristics of the Study's Theoretical Sample

Organizational Position	Age	Educational Level	Field and Degree	Relevant Work Experience (Years)	Interview Code
University faculty member	47	PhD	Accounting	18	M1
Provincial director general of tax affairs	52	Master's degree	Financial Management	25	M2
Financial manager of a publicly listed company	45	Master's degree	Accounting	20	M3
Senior researcher at the Tax Affairs Organization	41	PhD	Financial Economics	15	M4
Faculty member and tax consultant	50	PhD	Economics	22	M5
Head of the Tax Audit Department	48	Master's degree	Finance	23	M6
Tax manager of a large service company	43	Master's degree	Accounting	17	M7
Director of research and data analysis	39	PhD	Data Science	12	M8
Professional capital market investor	44	Master's degree	Financial Management	15	M9
Senior tax expert	40	Master's degree	Accounting	14	M10
Director of financial and tax affairs at a manufacturing company	46	Master's degree	Accounting	19	M11
Financial analyst and investment consultant	38	Master's degree	Financial Management	12	M12
Faculty member and tax law specialist	49	PhD	Financial Law	21	M13
Deputy director of the General Department of Tax Supervision	51	Master's degree	Public Administration	26	M14
Director of technology and tax data analytics	42	PhD	Information Technology	16	M15
Senior tax and accounting consultant	53	Master's degree	Accounting	28	M16
Financial manager of an engineering services company	39	Master's degree	Financial Management	13	M17
Senior risk and regulatory compliance specialist	37	Master's degree	Economics	11	M18

Director of the tax planning unit at a governmental organization	45	PhD	Public Policy	18	M19
Professional investor and market analyst	43	Master's degree	Financial Management	14	M20

3. Findings and Results

Step 1: Open Coding

This stage of the grounded theory method begins immediately after the first interview. In other words, following each interview, the researcher begins identifying concepts, assigning appropriate labels to them, and integrating related concepts. Open coding is the part of the analytical process that specifically involves naming and categorizing phenomena through their detailed examination (Strauss & Corbin, 1998). During the open-coding stage, data obtained from 20 interviews were analyzed line by line, resulting in the identification of 230 initial codes. These codes were extracted as basic themes related to data quality, information transparency, technological infrastructure, institutional integration, laws and regulations, tax evasion, investment efficiency, and corporate governance. The codes indicated that the country's current taxation system faces challenges such as information inconsistency, weak interconnectivity among government agencies, the absence of data standards, inadequate data-driven supervision, fictitious activities, inaccurate reporting, and insufficient technological infrastructure. Numerous codes also highlighted the role of big data in reducing concealment, increasing transparency, detecting irregularities, reducing tax evasion, improving financial reporting, increasing accountability, enhancing companies' ethical conduct, and creating information symmetry for investors. The extracted initial codes provided the foundation for organizing the data into categories and subcategories during the second stage, namely axial coding, and helped clarify the principal dimensions of the integrated big data-based tax management model.

Step 2: Axial Coding

In the second coding stage, known as axial coding, the researcher selects one category as the central category, examines it as the core phenomenon at the center of the process, and determines how the other categories are related to it (Strauss & Corbin, 1998). During the axial-coding stage, six main categories were identified after the aggregation of the 230 initial codes. These categories resulted from the repeated integration and comparison of codes and the examination of their causal relationships, contextual and intervening conditions, proposed strategies, and expected consequences. The results of axial coding indicated that:

1. Data quality and information transparency constitute the primary foundation for establishing a data-driven taxation system.
2. Technological infrastructure and intelligent systems enable the accurate analysis of taxpayers' behavior.
3. Institutional and legal factors play a decisive role in data coordination and exchange.
4. Tax evasion can be substantially identified through big data analytics.
5. Reducing investment inefficiency is a direct consequence of information transparency.
6. This structure ultimately provides the basis for selective coding and the presentation of the study's final model.

Step 3: Selective Coding

Selective coding involves systematically selecting the core category, relating it to the other categories, validating the relationships, and filling gaps with categories that require further refinement or development. This process consists of several steps:

The first step involves describing the main storyline. The second step involves relating the supplementary categories to the core category by using a paradigm model, as described during axial coding. The third step involves connecting the categories to one another at the next analytical level. The fourth step involves validating these relationships against the data. The final step involves completing categories that require refinement, elaboration, or further development (Naqavi, 2019, p. 104, as cited in Khosravi & Soroush). After formulating the proposed model based on the coded interview data, a questionnaire is designed. Questionnaires and measurement scales are probably the most widely used research instruments in the social sciences. Their low cost, limited resource requirements, and considerable potential for collecting data from large samples have made them effective and appealing research instruments for researchers and practitioners. As one of the most common data-collection instruments in survey research, a questionnaire consists of a set of goal-oriented questions that use various scales to measure a respondent's opinions, perspectives, and perceptions. A Likert scale, as presented in the following table, is used to score and assign values to the responses.

Selective coding involves systematically selecting the core category, relating it to the other categories, validating the relationships, and filling gaps with categories that require further refinement or development. This process consists of several steps:

The first step involves describing the main storyline. The second step involves relating the supplementary categories to the core category by using a paradigm model, as described during axial coding. The third step involves connecting the categories to one another at the next analytical level. The fourth step involves validating these relationships against the data. The final step involves completing categories that require refinement, elaboration, or further development (Naqavi, 2019, p. 104, as cited in Khosravi & Soroush).

During the selective-coding stage, the primary objective was to integrate the categories derived from axial coding and formulate the core phenomenon. This phenomenon organizes the other categories around itself and explains the network of meanings and causal and consequential relationships within a unified theoretical model. Following the analysis of 230 open codes and six axial categories, the central phenomenon capable of integrating all the identified concepts, strategies, conditions, and consequences into a coherent theoretical model was defined as follows:

Core phenomenon: Establishing an integrated big data-based tax collection and management model and its effect on improving companies' financial and social behavior

This core phenomenon demonstrates that big data are not merely a technological instrument; rather, they function as a driver of institutional and economic transformation and facilitate more accurate monitoring and appropriate corporate resource allocation through information transparency.

To clarify the process, the relationships among the categories are explained according to the principal components of the paradigm model, including causal conditions, contextual conditions, intervening conditions, strategies, and consequences.

Causal conditions: Conditions that lead to the emergence or formation of the core phenomenon:

1. Lack of data standardization and data inconsistency
2. Fictitious reporting and the provision of inaccurate corporate information
3. Absence of a national integrated tax database
4. Weak technological infrastructure and the absence of intelligent analytical systems
5. Insufficiently transparent regulations governing interorganizational data exchange

Collectively, these factors justify and necessitate the development of an integrated big data-based taxation model.

Contextual conditions: Environmental conditions that facilitate or necessitate the implementation of the phenomenon:

1. The increasing volume of banking and financial transactions
2. The growth of the digital economy and the expansion of big data
3. Pressure from regulatory bodies and the media to improve financial transparency
4. The capital market's growing demand for accurate and reliable information
5. National macro-level policies aimed at combating corruption and tax evasion

These contextual factors make the establishment of a data-driven taxation model a national necessity.

Intervening conditions: Factors that positively or negatively affect the implementation of the model:

1. Organizational resistance to data sharing
2. A shortage of human resources specializing in data analytics
3. Data security risks and cyber threats
4. Differences among organizations in interpreting tax regulations
5. A weak organizational culture regarding data-driven operations

These factors operate as filters that may either accelerate or impede the implementation and effectiveness of the model.

Strategies: Measures adopted to realize the core phenomenon:

1. Establishing a centralized national platform for integrating tax and economic data
2. Developing machine-learning algorithms to detect tax-evasion patterns
3. Standardizing data across banks, insurance companies, customs authorities, and the Tax Affairs Organization
4. Designing intelligent decision-support dashboards for managers
5. Improving information security and data encryption
6. Reforming the laws and regulations governing interagency data exchange
7. Strengthening internal controls and transparent corporate reporting systems

These strategies provide the pathway for the full implementation of the integrated taxation model.

Consequences: The outcomes of implementing the core phenomenon, which were explicitly extracted from the interviews:

Tax-related consequences

- A substantial reduction in tax evasion
- Increased accuracy in identifying high-risk taxpayers
- Increased government tax revenue
- Improved financial integrity and the prevention of corruption

Economic consequences

- Reduction in fictitious and inefficient investments
- Direction of capital toward financially sound and productive companies
- Improved resource-allocation efficiency in the capital market

Table 2. Results of the Qualitative Analysis

Paradigm Model Dimension	Content Description	Examples of Extracted Codes
Causal conditions	Factors that lead to the formation of, or create the need for, the core phenomenon	Data inconsistency, fictitious reporting, weak information technology infrastructure, lack of standardization, absence of transparent regulations, and inconsistency between tax returns and actual circumstances
Contextual conditions	Environmental conditions that facilitate or necessitate the implementation of the phenomenon	Growth of big data, increased transaction volumes, media pressure, the capital market's demand for transparency, and anti-corruption policies
Intervening conditions	Factors that strengthen or weaken the implementation of the strategies	Organizational resistance, shortage of specialists, data security risks, differing interpretations of regulations, weak internal controls, and the absence of a data-driven culture
Strategies	Actions and policies adopted to realize the phenomenon	Establishment of a national data platform, standardization, data integration, tax-evasion detection algorithms, intelligent dashboards, encryption, and legal reforms
Consequences	Outcomes of realizing the phenomenon at the economic, social, and organizational levels	Reduced tax evasion, increased government revenue, reduced investment inefficiency, improved capital efficiency, enhanced corporate social responsibility, and improved corporate governance

In this qualitative study, data were collected through interviews with experts and specialists in tax management and big data. After transcription, the interviews were carefully reviewed and summarized several times, and statements related to big data-based tax collection and management and their effects on corporate investment inefficiency were extracted. Data analysis was based on Strauss and Corbin's grounded theory approach and consisted of three stages: open, axial, and selective coding. During open coding, key concepts were extracted from the interview transcripts and summarized using conceptual codes, such as data quality, the absence of data standards, and information heterogeneity. This process made it possible to transform extensive textual data into an analyzable set of concepts. One of the principal themes emerging from the findings was data quality and standardization. The experts emphasized the absence of common standards and the poor quality of data, which create analytical problems, result in inappropriate decision-making, and reduce the efficiency of the taxation system. Institutional integration and information exchange were also identified as key factors in reducing information gaps and preventing financial misconduct. The role of technology and data analytics in intelligent supervision constituted another major theme. The experts believed that data-analytics tools could identify violations more rapidly and reduce reliance on manual examinations. Furthermore, concepts such as information transparency, data security, corporate governance, and internal control were highlighted as factors that improve tax-related behavior and enhance the efficiency of the supervisory system. Ultimately, the analysis of the interviews indicated that developing a data-driven taxation system requires simultaneous attention to structural, technological, and legal dimensions. Data standardization, strengthened interorganizational cooperation, the development of data-analytics capabilities, increased information transparency and security, and legal reforms are among the prerequisites for establishing an efficient and effective data-driven system for financial supervision.

4. Discussion and Conclusion

The present study developed an integrated model for big data-based tax collection and management and examined its implications for corporate investment inefficiency. The qualitative findings showed that the establishment of such a system depends on the interaction of several structural, technological, legal, institutional, and behavioral factors. The central phenomenon identified in the selective-coding stage was the establishment of

an integrated big data-based tax collection and management model capable of improving the financial and social behavior of companies. The findings further indicated that poor data quality, lack of standardization, fragmented databases, fictitious reporting, weak technological infrastructure, and insufficiently transparent data-exchange regulations constitute the principal causal conditions that create the need for a data-driven taxation system. This result is consistent with previous studies emphasizing that electronic tax collection cannot be effectively implemented without integrated information systems, standardized data, and coordinated administrative structures [4-6].

One of the most important findings was the central role of data quality and standardization in the successful operation of the integrated tax system. The interviewees repeatedly emphasized that inaccurate, incomplete, inconsistent, and nonstandardized data weaken the reliability of tax analyses and may lead to incorrect assessments of taxpayers and companies. This finding suggests that the effectiveness of big data is not determined merely by the volume of available information. Rather, the usefulness of big data depends on whether data from banks, insurance companies, customs authorities, corporations, and governmental agencies are comparable and capable of being integrated. Previous research has similarly shown that big data analytics can improve tax collection efficiency only when administrative data are sufficiently reliable and structured for risk assessment and taxpayer classification [3]. The conceptual models developed for electronic tax collection also identify information quality, interoperability, technological infrastructure, and systematic integration as foundational components of effective tax administration [4, 5].

The findings also showed that institutional integration and interorganizational data exchange are indispensable for reducing information gaps and detecting tax evasion. The participants believed that the fragmentation of information across multiple organizations enables taxpayers and companies to provide inconsistent reports without immediate detection. An integrated national platform could allow the tax administration to compare tax declarations with banking records, customs data, insurance information, ownership records, and corporate financial statements. This result aligns with the comprehensive electronic tax collection model proposed by Rabiei et al., which emphasizes the interdependence of administrative coordination, information exchange, technological integration, and regulatory support in reducing tax evasion [8]. It is also compatible with findings concerning the organizational complexities of tax collection in Iran, where weak coordination, procedural fragmentation, and the multiplicity of involved institutions have been identified as major barriers to effective tax administration [6]. Therefore, the present study confirms that technological integration must be accompanied by institutional integration.

Another major finding concerned the role of intelligent technologies in transforming tax supervision from a reactive process into a preventive and risk-based process. The experts emphasized that machine learning, automated cross-checking, intelligent dashboards, and predictive analytics can identify suspicious patterns more rapidly than conventional manual audits. Such technologies can classify taxpayers according to risk, detect irregular transactions, identify inconsistencies between declarations and actual economic activities, and allocate audit resources more efficiently. This finding is supported by research demonstrating that artificial intelligence can optimize tax compliance, fraud detection, and revenue collection through predictive modeling and automated pattern recognition [10]. The proposed use of self-organizing maps for value-added tax collection also demonstrates that artificial intelligence can distinguish normal and anomalous taxpayer behavior and support the identification of potentially fraudulent cases [11]. Moreover, the integration of blockchain and artificial intelligence can strengthen transaction traceability and reduce the manipulation of tax information [12].

The results indicated that intelligent tax monitoring can make a substantial contribution to reducing tax evasion. Participants described tax evasion as being associated with fictitious invoices, inaccurate declarations, hidden transactions, artificial expenses, and inconsistencies between reported and actual financial activities. By integrating multiple sources of data, tax authorities can reduce the opportunity to conceal economic activities and increase the probability that irregularities will be detected. This finding is consistent with studies showing that smart governance and electronic tax systems can strengthen the control of tax evasion by improving transparency, increasing the speed of information access, and enabling systematic verification of taxpayer information [8, 9]. It also supports the argument that tax audits, compliance supervision, and effective collection mechanisms contribute to increased tax revenue when they are based on reliable information and targeted enforcement [13]. Thus, big data analytics can enhance the effectiveness of conventional enforcement instruments rather than completely replacing them.

The findings also revealed that information transparency is a key mechanism linking big data-based taxation to corporate investment efficiency. The experts believed that the integration of financial and tax information reduces information asymmetry among managers, investors, creditors, regulators, and tax authorities. When companies are aware that their transactions and declarations can be cross-checked against independent databases, they may be less likely to engage in fictitious investment, conceal resources, manipulate expenses, or allocate funds to projects that primarily serve managerial interests. This result suggests that an integrated tax system can influence corporate behavior beyond the direct objective of tax collection. The finding is compatible with research emphasizing that transparent and fair tax systems strengthen accountability and reduce opportunities for opportunistic conduct [2, 16]. The development of electronic tax collection models has similarly highlighted transparency, accountability, and information integrity as necessary conditions for effective fiscal governance [5].

The reduction of investment inefficiency identified in the present study can also be explained through improvements in corporate governance and internal control. The interviewees argued that the traceability of transactions and the integration of tax data increase managerial accountability and reduce the possibility of using corporate resources for nonproductive or concealed activities. Stronger tax information systems may also help boards of directors, auditors, and investors assess the actual economic condition of firms. In this regard, the findings show that big data-based tax administration can act as an external governance mechanism that complements internal controls. The emphasis placed by previous studies on fair tax collection, good governance, and appropriate enforcement supports this interpretation because transparent procedures and predictable oversight can improve both institutional legitimacy and corporate compliance [2, 16]. In addition, tax incentives and comprehensive tax planning can affect companies' financial behavior and resource-allocation decisions, indicating that the design of taxation systems has broader consequences for corporate investment strategies [17].

Data security emerged as another essential component of the model. The participants emphasized that the integration of large volumes of financial, banking, and corporate information creates significant risks related to unauthorized access, cyberattacks, misuse of information, and breaches of confidentiality. These risks can reduce institutional willingness to share data and weaken taxpayers' trust in digital taxation systems. Therefore, the implementation of encryption, access controls, secure storage, authentication protocols, and clear rules concerning data ownership was identified as necessary. The importance of security is consistent with studies of blockchain-based tax administration, which suggest that tamper-resistant and traceable records can strengthen the integrity of tax information [12]. It is also related to the broader requirement that electronic tax systems be designed around

both technological capability and taxpayer trust. When taxpayers perceive digital systems as secure, transparent, and reliable, their satisfaction and willingness to use electronic tax services may increase [14].

The study further identified organizational resistance, shortage of specialized personnel, inconsistent interpretations of tax law, and weak data-driven culture as intervening conditions that may hinder implementation. This finding demonstrates that technological systems cannot produce the expected outcomes without organizational readiness. Employees must possess the analytical skills required to interpret complex data, managers must support evidence-based decision-making, and institutions must be willing to modify established routines. The organizational challenges identified in previous tax research support this result, particularly the evidence that administrative complexity, resistance, procedural weakness, and coordination problems can reduce the effectiveness of tax collection reforms [6]. The digital divide may also appear within public organizations, because unequal access to technological knowledge and infrastructure can prevent some departments or regions from benefiting equally from data-driven systems [15]. Therefore, capacity building and organizational change management are essential dimensions of the proposed model.

The findings regarding taxpayer satisfaction and fairness indicate that the legitimacy of the integrated system will depend on how taxpayers experience it. The participants suggested that a data-driven system should not merely increase surveillance and penalties but should also simplify procedures, reduce repetitive documentation, accelerate tax assessments, and provide clearer information to taxpayers. Electronic tax collection has been shown to influence taxpayer satisfaction, particularly when digital systems are accessible, responsive, and easy to use [14]. Moreover, fair treatment, transparent assessment criteria, and proportionate sanctions are important for maintaining trust in tax administration [2, 16]. If big data systems are perceived as arbitrary, intrusive, or error-prone, they may generate resistance rather than compliance. Accordingly, the proposed model should balance enforcement efficiency with procedural fairness and taxpayer rights.

The economic consequences identified in the study included increased tax revenue, improved allocation of capital, reduced fictitious investment, and stronger financial integrity. These consequences indicate that effective tax collection can support not only government finances but also broader social and economic development. Research has shown that tax revenue transfers can contribute to regional social development and the financing of public services [18]. From this perspective, reducing tax evasion expands the resources available for public investment and decreases the unfair burden imposed on compliant taxpayers. Waris has similarly argued that financial resource collection and allocation should be connected to institutional fairness, accountability, and the realization of social rights [1]. Thus, the proposed model has implications beyond administrative efficiency, as it may strengthen the fiscal capacity of government and improve the legitimacy of public financial institutions.

The results also confirm that the core phenomenon should be understood as an institutional transformation rather than a narrowly technological reform. Big data technologies can facilitate data analysis, but the effectiveness of the model depends on laws, organizational structures, governance mechanisms, human competencies, and public trust. This interpretation is aligned with previous models of electronic tax collection that combine technological, organizational, legal, managerial, and environmental dimensions [4, 5, 8]. The findings also reinforce the importance of designing an integrated information architecture that is coherent and understandable for both officials and taxpayers. Although Turgay and Ünlü examined the memorability and comprehensibility of physical spaces, their emphasis on coherent structural organization can be conceptually extended to the design of complex tax information systems, in which clear pathways and recognizable structures facilitate effective use [19].

Overall, the study extends the existing literature by connecting big data–based tax administration with corporate investment inefficiency. Previous studies have mainly examined tax collection, tax evasion, taxpayer compliance, electronic systems, artificial intelligence, and organizational challenges as separate issues. The present findings integrate these elements within a single grounded theoretical model and show how causal conditions, contextual factors, intervening conditions, strategies, and consequences interact. The proposed model suggests that standardized and integrated data, intelligent analytics, institutional coordination, legal reform, information security, and strong internal controls jointly improve tax transparency and reduce opportunities for inefficient corporate investment. This integrated interpretation is consistent with the broader literature on smart taxation, fair collection, artificial intelligence, and good fiscal governance [2, 9, 10, 16].

This study was subject to several limitations. First, the qualitative design and the use of purposive expert sampling limit the statistical generalizability of the findings. Although the participants had substantial experience in taxation, finance, data analytics, and investment, their views may not fully represent all stakeholders, particularly taxpayers, small business owners, software developers, auditors, and policymakers from different regions. Second, some participants may have been cautious when discussing organizational weaknesses, tax evasion, and data-sharing problems because of the sensitivity of these subjects. Third, the proposed model was derived from interview data and was not tested using actual tax records, corporate investment data, or longitudinal observations. Finally, the rapid development of artificial intelligence, cybersecurity, and digital taxation may change some technological and institutional requirements over time.

Future studies should quantitatively validate the proposed model using structural equation modeling and data collected from larger and more diverse samples. Researchers should examine the direct and indirect relationships among data quality, institutional integration, technological readiness, tax transparency, corporate governance, tax evasion, and investment efficiency. Longitudinal studies could evaluate whether the implementation of integrated tax platforms produces sustained changes in tax compliance and corporate investment behavior. Comparative research across countries, provinces, industries, and firm sizes would also help determine whether the model performs differently under varying institutional and technological conditions. Future studies may further examine ethical concerns, algorithmic bias, taxpayer privacy, cybersecurity risks, and the consequences of automated tax decisions.

Tax authorities should establish a centralized and secure national data platform capable of integrating information from banks, customs agencies, insurance organizations, corporate registries, municipalities, and other relevant institutions. Uniform data standards and clear legal protocols for data exchange should be developed before large-scale analytical systems are implemented. Tax administrations should invest in machine-learning tools, intelligent audit dashboards, cybersecurity infrastructure, and specialist training for employees. Companies should also strengthen internal controls, improve the accuracy of financial reporting, and establish clear responsibilities for tax data management. Finally, implementation should be gradual, transparent, and accompanied by taxpayer education, appeal mechanisms, independent oversight, and continuous evaluation of system accuracy and fairness.

Authors' Contributions

Authors equally contributed to this article.

Ethical Considerations

All procedures performed in this study were under the ethical standards.

Acknowledgments

Authors thank all participants who participate in this study.

Conflict of Interest

The authors report no conflict of interest.

Funding/Financial Support

According to the authors, this article has no financial support.

References

- [1] A. Waris, "A Principled Approach to Financial Resource Collection and Allocation for the Realization of Human Rights in Creating a Global Fiscal Institution," *Bull. Intl. Taxn.*, vol. 78, no. 9/10, 2024, doi: 10.59403/36t1v90.
- [2] R. Ameri, S. J. Mir Ghasemi, and V. Rostami, "An Optimal Model for Tax Collection Sanctions under Principles of Good Governance," *Journal of New Administrative Law Studies*, no. 9, 2021.
- [3] J. Atanasijević, D. Jakovetic, N. Krejić, N. Krklec Jerinkic, and D. Marković, "Using big data analytics to improve efficiency of tax collection in the tax administration of the Republic of Serbia," *Ekonomika Preduzeća*, p. 115, 2018, doi: 10.5937/EkoPre1808115A.
- [4] F. Fathi, F. Namian, A. Baghlaki, and S. R. Hosseini, "Identification, Integration, Weighting, and Presentation of a Conceptual Model of Factors Affecting Electronic Tax Collection Using Meta-Analysis and Shannon Entropy Method," *Advertising and Sales Management*, no. Online Publication, pp. 316-333, 2023. [Online]. Available: https://asm.pgu.ac.ir/article_703332.html.
- [5] F. Fathi, F. Namamian, A. Baghfalaki, and S. R. Hassani, "Identification, Integration, Weighting, and Presentation of a Conceptual Model of Components Affecting Electronic Tax Collection Using Meta-Synthesis and Shannon Entropy," *Advertising and Sales Management*, pp. 316-333, 2024.
- [6] H. Ghasemi Bonabari and E. Goudarzi, "Organizational Challenges and Management of Tax Collection Complexities in Iran," (in English), *Public Service Management Studies*, vol. 3, no. 7, pp. 205-242, 2025, doi: 10.22054/spsa.2025.89169.1078.
- [7] R. Mu, N. M. Fentaw, and L. Zhang, "Tax Evasion, Psychological Egoism, and Revenue Collection Performance: Evidence from Amhara Region, Ethiopia," *Frontiers in Psychology*, vol. 14, p. 1045537, 2023, doi: 10.3389/fpsyg.2023.1045537.
- [8] K. Rabiei, M. Fazeli, and A. Ahmadi Khoshabari, "Presenting a Comprehensive Electronic Tax Collection Model to Reduce Tax Evasion Using the ISM Approach," (in English), *Governmental Accounting*, vol. 10, no. 2, 2024.
- [9] A. Ashrafi, "Tax Collection in Smart Governance and Its Role in Controlling Tax Evasion," *23rd International Conference on Research in Management, Economics, and Development*, 2024. [Online]. Available: <https://civilica.com/doc/2220871/>.
- [10] S. Rahman, M. R. M. Sirazy, R. Das, and R. S. Khan, "An exploration of artificial intelligence techniques for optimizing tax compliance, fraud detection, and revenue collection in modern tax administrations," *International Journal of Business Intelligence and Big Data Analytics*, vol. 7, no. 3, pp. 56-80, 2024.
- [11] F. O. Bankole and Z. Vara, "Artificial Intelligence Systems for Value Added Tax Collection via Self Organizing Map (SOM)," 2022, doi: 10.21203/rs.3.rs-2216885/v1.
- [12] M. Gonzalez and P. Alvarez, "Blockchain and Artificial Intelligence: Enhancing Tax Collection in Spain," *International Journal of Taxation and Digital Technologies*, vol. 10, no. 1, pp. 89-102, 2023.
- [13] K. Kamaruddin, A. Faisal, and M. R. Agustam, "The Effect of Tax Audit, Tax Collection, and Taxpayer Compliance Supervision on Increasing Tax Revenue at KPP Pratama Makassar Utara," *Golden Ratio of Taxation Studies*, vol. 2, no. 1, pp. 36-46, 2022, doi: 10.52970/grts.v2i1.245.
- [14] F. Tahmasbi, "The Effect of Electronic Tax Collection on Taxpayer Satisfaction: A Case Study of Taxpayers in Qazvin," in *Fourth National Conference on Interdisciplinary Research in Management and Humanities*, Tehran, 2023.
- [15] S. Torkashvand, B. Yarigholi, and V. Moradiyan Mohammadi, "Explanation of the challenges of the digital divide over the implementation of educational justice," (in en), *Technology of Education Journal (TEJ)*, vol. 16, no. 2, pp. 263-280, 2022, doi: 10.22061/tej.2021.7363.2527.

- [16] K. Marahami, A. Mohammadzadeh, N. Hamidi, and M. Kheirandish, "Presenting a Model for Fair Tax Collection in the Country," (in English), *Journal of Development and Transformation Management*, vol. 16, no. 57, 2024.
- [17] S. Hassanzadeh, Y. Kamyabi, M. Khalilpour, and J. Ramzani, "The Effect of Tax Incentives on Tax Collection in the Comprehensive Tax Plan Using Grounded Theory and Structural Equation Model," *International Journal of Nonlinear Analysis and Applications*, vol. 14, no. 5, pp. 113-120, 2023, doi: 10.22075/ijnaa.2022.27266.3544.
- [18] C. O. Omodero, "Value-Added Tax Revenue Transfers and Regional Social Development: Evidence from Nigeria," *Sustainability*, vol. 14, no. 21, doi: 10.3390/su142114343.
- [19] Z. T. Turgay and A. Ünlü, "What Makes a Space Relatively Memorable a Study on the Recollection of Spaces Through Space Syntax and Imageability Theories," *Iconarp International J of Architecture and Planning*, 2025, doi: 10.15320/iconarp.2025.322.