

INNOVATION SCIENCE AND TECHNOLOGY



Scopus || Electronic journal specializing in Scopus

ISSUE 1



Acceptance of papers **January, 2026**



**Acceptance of
papers**

Published monthly



Topics

economics,
technology, social
sciences

ISSN 3060-5229



Digital
Object
Identifier



Visit the website
t.me/scopus_IST2100

Editorial board:



EDITOR-IN-CHIEF:

Mirzaliyev Sanjar Makhmatjon ugli

DEPUTY EDITOR-IN-CHIEF:

Makhmudov Nosir Makhmudovich
DSc., Prof., Academician

DEPUTY EDITOR-IN-CHIEF:

Ochilov Bobur Bakhtiyor ugli – Senior
lecturer at TSUI

THE SCIENTIFIC-POPULAR ELECTRONIC
JOURNAL **"INNOVATION SCIENCE AND
TECHNOLOGY"** HAS BEEN REGISTERED
UNDER THE NUMBER **C-5669633** BY THE
AGENCY FOR INFORMATION AND MASS
COMMUNICATIONS (AOKA) OF THE
REPUBLIC OF UZBEKISTAN, EFFECTIVE
FROM OCTOBER 9, 2024.

CONTACTS

Phone: **+998 50 737 87 88**

Website: <https://ist-journal.uz>

Email: innovationist2025@gmail.com

The scientific electronic journal "Innovation Science and Technology" has been included in the list of scientific publications recommended for the publication of main scientific results of dissertations for the award of PhD and DSc degrees in economics and technical sciences, in accordance with the Resolution No. 370 of the Presidium of the Higher Attestation Commission of the Republic of Uzbekistan, dated May 8, 2025.



Sharipov Kongiratbay Avezimbetovich,
Doctor of Technical Sciences (DSc), Professor



Abdurakhmanova Gulnora Kalandarovna, Doctor of
Economic Sciences (DSc), Professor



Cham Tat Huei,
Doctor of Philosophy (PhD), Professor (Malaysia)



Muhammad Imran Sadiq
Doctor of Philosophy in Economics (PhD), Professor,
Malaysia



Ahmed Aziz Ismail
Doctor of Technical Sciences (DSc),
Professor (Egypt)



Lee Chin
Doctor of Philosophy in Economics (PhD), (Malaysia)



Asongu SImplice
Doctor of Philosophy in Economics (PhD), Cameroon



Rui Dang
Doctor of Chemistry (DSc), Professor, China



Zahoor Ahmed
Doctor of Philosophy in Economics (PhD), Turkey



Shujaat Abbas
Doctor of Philosophy in Economics (PhD), Russia



Tina A Coffelt
Doctor of Philosophy in Educational Sciences (PhD),
USA



Abdikarimova Dinara Rustamxanovna
Doctor of Economic Sciences (DSc), Professor

Kurbonbekova Mohichehra Turobjonovna
Doctor of Economic Sciences (DSc), Professor

Alimardonov Ilkhom Muzrabshokovich
Doctor of Economic Sciences (DSc), Professor



Razakova Barno Sayfiyevna
Doctor of Philosophy in Economics (PhD)



Khasanov Sarvar Ulugbek ugli
Doctor of Philosophy in Economics (PhD)



Kholikova Rukhsora Sanjarovna
Associate Professor (PhD)

CONTENTS

BRIEF FEEDBACK ON “AGAT CREDIT” MICROFINANCE ORGANIZATION BASED ON THE REPORT OF “KAPDEPO” INVESTMENT COMPANY: CAVEATS FOR LENDERS (BONDHOLDERS)	16
Abduganiev Abdulaziz Alisher ugli	
IMPLEMENTATION OF EU BEST AGRICULTURAL TRADE PRACTICES IN UZBEKISTAN.....	20
Khulkar Karimova Rakhmanali qizi	
THE ROLE OF FOREIGN DIRECT INVESTMENT IN INCREASING SERVICE EXPORTS OF UZBEKISTAN	26
Jamshid Mirzakhmedov	
THE ROLE AND IMPORTANCE OF FINANCIAL MARKETS IN ECONOMIC DEVELOPMENT	30
Baymanova Mavlyuda Djuraevna, Abdullaeva Shohista, Ubaydullaeva Gulchehra Erkabaevna	
КЛИНИЧЕСКАЯ ОЦЕНКА СОСТОЯНИЯ МЕСТНЫХ ИММУННЫХ МЕХАНИЗМОВ ПОЛОСТИ РТА У ПАЦИЕНТОВ НА ЭТАПАХ ОРТОДОНТИЧЕСКОГО ЛЕЧЕНИЯ	36
Рахимбердыев Рустам Абдунасирович, Сайфулаева Азиза Анваровна	
INTEGRATING AI-BASED CUSTOMER ANALYTICS INTO INNOVATIVE RETAIL MARKETING STRATEGIES	40
Ostonaqulova Gulsaraxon Muhammadyoqub qizi	
FINANCIAL STIMULATION OF INNOVATIVE ACTIVITIES OF ENTERPRISES THROUGH INVESTMENTS	48
Bahriddinov Nodirbek Zamirdinovich	
DIGITAL DENTISTRY: LITERATURE REVIEW	52
Tursunov Begzod Sherzodovich, Zokirova Nodira Sobitovna	
THE LATEST ADHESIVE TECHNOLOGIES IN DENTISTRY	56
Rahimberdiyev Rustam Abdunasirovich, Chinibayeva Ibagul Sarsenbayevna	
ENSURING THE ACCEPTABILITY OF QUANTITATIVE AND QUALITATIVE INDICATORS IN THE EFFECTIVE ORGANIZATION OF HOUSING FUNDS IN KHOREZM	61
Otajonov Tohirjon Khojanazar o'g'li	
WAYS TO IMPROVE CUSTOMS ADMINISTRATION IN THE REPUBLIC OF UZBEKISTAN.....	67
Usmonova Dilfuza Ilhomovna	
CLINICAL ASSESSMENT OF THE STATE OF LOCAL IMMUNE MECHANISMS OF THE ORAL CAVITY IN PATIENTS AT DIFFERENT STAGES OF ORTHODONTIC TREATMENT	72
Rakhimberdiyev Rustam Abdunasirovich, Saifulaeva Aziza Anvarovna	
IMPROVING THE ALGORITHM FOR CONTROLLING THE CUSTOMS TRANSIT INFORMATION SYSTEM E-TRANSIT OF THE REPUBLIC OF UZBEKISTAN	76
Musayeva Shoirazimovna	
DEVELOPMENT TRENDS OF THE AUTOMOTIVE BUSINESS IN UZBEKISTAN	82
Saidov Dilshodbek Razzakovich	
INTEGRATION OF MARKETING STRATEGIES IN RETAIL TRADE ACTIVITIES.....	87
Akramov Toxir Abdiraxmanovich	
CHALLENGES OF ADOPTING ISLAMIC FINANCE WITHIN CONVENTIONAL BANKING SYSTEMS	91
Safarov Shuhrat Ismatovich	
CRM SYSTEMS AND THEIR IMPACT ON THE RESULTS OF MARKETING STRATEGY IN DISTRIBUTION COMPANIES	95
Jamoliddinov Fakhriyor Shodiyor o'g'li	
LEXICAL-SEMANTIC ARCHITECTURE OF MODERN WORDNET SYSTEMS	101
Aynura Axmedova	
METHODOLOGY FOR ANALYZING THE EFFECTIVENESS OF INNOVATIVE PROCESSES AT ENTERPRISES.....	108
Kurbanova Shakhnoza Yuldashbayevna	
COMPANY VALUATION IN MERGERS AND ACQUISITIONS: A STRATEGIC AND GOVERNANCE-BASED APPROACH	113
Lee Illarion Georgievich	

A REVIEW OF THE LITERATURE ON CAD/CAM TECHNOLOGIES IN DENTAL ECTOPROSTHETICS.....	118
Tursunov Begzod Sherzodovich, Hazratqulov Asrbek Ulugbek ugli	
TRENDS AND DIFFICULTIES IN THE INTEGRATION OF DIGITAL TECHNOLOGIES IN ORTHOPEDIC DENTISTRY.....	123
Khojimurodov Burkxon Ravshanovich	
PRIORITY DIRECTIONS FOR IMPROVING THE MECHANISM OF ENHANCING THE ECONOMIC SECURITY LEVEL OF THE KASHKADARYA REGION.....	127
Tuyev Abdurahmon Yusubovich	
THE ROLE OF PSYCHOPHYSIOLOGICAL TRAINING OF DRIVERS IN REDUCING ROAD TRAFFIC ACCIDENTS.....	132
Uralbayev Anvar Ubaydullayevich	
THE ROLE OF SUSTAINABLE DEVELOPMENT PRINCIPLES IN DEVELOPING GREEN MARKETING STRATEGIES FOR ENTERPRISES.....	135
Sapayev Akhmad Durdibayevich	
MANAGEMENT MODEL OF INFORMATION RESOURCES IN SMALL BUSINESS ENTITIES AND ITS IMPACT ON ECONOMIC EFFICIENCY.....	140
Yo'ldoshev Nodirbek Ne'matjon o'g'li	
WAYS TO DEVELOP THE INVESTMENT ACTIVITY OF COMMERCIAL BANKS THROUGH THE SECURITIES MARKET.....	145
Yuldashev Fozil Turapovich	
INTERNATIONAL EXPERIENCE IN THE USE OF CROSS-BORDER REMITTANCES IN THE DEVELOPMENT OF THE NATIONAL ECONOMY.....	152
Gimranova O. B.	
FREE ECONOMIC ZONES AND FOREIGN INVESTMENT.....	158
Sheraliyeva Saida Azatovna	
ISSUES OF FORMATION AND MANAGEMENT OF PRODUCT ASSORTMENT IN RETAIL ENTERPRISES.....	162
Safarov Baxtiyor Djurakulovich	
STATE SUPPORT IN THE REPUBLIC OF UZBEKISTAN FOR ORGANIZING SHORT-TERM SCIENTIFIC INTERNSHIPS OF YOUNG SCIENTISTS ABROAD.....	167
Kabashev Tairjon	
LEGAL FOUNDATIONS OF DIVIDEND POLICY: EVIDENCE FROM DEVELOPED AND DEVELOPING COUNTRIES.....	172
Eshev Furqat A'zamovich	
IMPROVING SMART CITY GOVERNANCE BASED ON DIGITAL PLATFORMS: A HUMAN-CENTERED APPROACH.....	176
Rakhimova Madina Shukhrat qizi	
THE INVESTMENT CLIMATE AND ITS IMPACT ON THE DEVELOPMENT OF FOREIGN TRADE: A CASE STUDY OF UZBEKISTAN.....	182
Mirzamukhamedova Shakhzoda Akmaljon qizi	
CONSUMER CREDITS IN USA.....	187
Zunnunova Xulkar Muxtorovna	
INSTITUTIONAL BASES AND FUNCTIONAL MECHANISMS OF CONTROLLING IN THE EFFECTIVE MANAGEMENT OF THE RAILWAY TRANSPORT SYSTEM.....	194
Kayumov Zafarbek Odil ugli	
ANALYSIS OF CUSTOMER RELATIONSHIP MANAGEMENT PROCESSES AND PROBLEMS IN SERVICE ENTERPRISES.....	199
Ismailova Ma'mura Eldorovna	
PROBLEMS FACED BY COMMERCIAL BANKS IN BANK RISK MANAGEMENT AND WAYS TO ADDRESS THEM.....	205
Qayimova Ismigul Ilhom qizi, Tuxsanov Eldor Dilmurod o'g'li	

DESIGN OF ENGINEERING STRUCTURES AND CONSTRUCTION OF A REGIONALLY BRANCHED HIGHWAY COMPLEX.....	209
Yakubov Maqsadkhon Sultaniyazovich, Norinov Muhammadyunus Usibjonovich, Zikraev Akmaljon Alimovich	
THE ROLE OF COOPERATIVE RELATIONS IN THE SUSTAINABLE DEVELOPMENT OF THE REGIONAL TOURISM MARKET	216
Mirzabayev Jamshid Irkinovich	
THE ROLE OF FREE ECONOMIC ZONES IN IMPROVING THE INVESTMENT CLIMATE OF THE KHOREZM REGION.....	221
Masharipov Sardorbek Farxadovich	
ANALYSIS OF ORGANIZATIONAL AND ECONOMIC INDICATORS OF INNOVATIVE POTENTIAL MANAGEMENT IN TEXTILE INDUSTRY ENTERPRISES	228
Khosilov Shavkat Bekmurodovich	
MAIN WAYS TO DEVELOP INTEREST RATE RISK MANAGEMENT PRACTICES IN COMMERCIAL BANKS OF UZBEKISTAN	234
Seytnazarov Daniyar Baxadirovich	
THEORETICAL AND METHODOLOGICAL ASPECTS OF ESG STRATEGY IMPLEMENTATION.....	239
Xusenova Mexrangiz	
ADVANTAGES OF USING TRADITIONAL CONSTRUCTION MATERIALS IN THE CONSTRUCTION OF LOW-RISE RESIDENTIAL BUILDINGS.....	244
Otabek Hakimovich Toshniyozov	
THE METHODOLOGY FOR SELECTING AND INTEGRATING DATA SOURCES AND USING OFFICIAL STATISTICAL ENTERPRISE DATA, QUESTIONNAIRES, AND PROXY INDICATORS IN FORMING THE EMPIRICAL BASIS OF THE STUDY.....	247
Usmonov Maxsud Tulqin o'g'li, Qodirov Farrux Ergash o'g'li	

THE METHODOLOGY FOR SELECTING AND INTEGRATING DATA SOURCES AND USING OFFICIAL STATISTICAL ENTERPRISE DATA, QUESTIONNAIRES, AND PROXY INDICATORS IN FORMING THE EMPIRICAL BASIS OF THE STUDY

Usmonov Maxsud Tulqin o'g'li

E-mail: maqsudu32@gmail.com

ORCID: 0000-0001-9997-6617

Scientific advisor:

Qodirov Farrux Ergash o'g'li

Shahrisabz State Pedagogical Institute

Mathematics and information technology in education

Head of the Department, ifd DSc, Associate Professor

E-mail: farrukhbek0209@mail.ru

ORCID: 0000-0002-4574-7728

Abstract: This paper develops a methodology for constructing an empirical research database through the systematic selection and integration of multiple data sources, including official statistics, firm-level data (accounting, managerial, transactional, and registry records), survey data, and proxy indicators for hard-to-measure constructs. A practical set of criteria for data source selection is proposed, covering representativeness, accuracy, coverage, temporal and spatial comparability, data preparation costs, legal and ethical constraints, and reproducibility. The data integration workflow is described step by step, encompassing identifier standardization and record linkage, harmonization of measurement units and classifications, panel data construction, metadata management, and version control. In addition, the study outlines key data quality checks, such as missing data diagnostics, outlier detection, internal consistency analysis, selection bias, and linkage bias, alongside documentation standards required for transparent and robust econometric research. The Results section presents illustrative tables and conceptual figures, including a data source matrix, an integration map, a proxy indicator catalogue, and a data quality scorecard. The paper concludes with practical recommendations for building reliable, auditable, and replicable empirical databases in applied economic research.

Key words: empirical database; data sources; official statistics; firm-level data; surveys; data integration; record linkage; proxy indicators; data quality; metadata.

Annotatsiya: Mazkur maqolada empirik tadqiqot bazasini shakllantirish uchun rasmiy statistika ma'lumotlari, korxonalar darajasidagi ma'lumotlar (buxgalteriya, boshqaruv, tranzaksiya va reyestr yozuvlari), so'rovnoma natijalari hamda bevosita o'lchash qiyin bo'lgan tushunchalar uchun proksi-indikatorlarni tanlash va integratsiyalash metodologiyasi ishlab chiqiladi. Ma'lumotlar manbalarini tanlash uchun amaliy mezonlar majmui taklif etiladi, jumladan: reprezentativlik, aniqlik, qamrov, vaqt va makon bo'yicha taqqoslanish imkoniyati, ma'lumotlarni tayyorlash xarajatlari, huquqiy va etik cheklorlar hamda takrorlanuvchanlik. Ma'lumotlarni integratsiyalash jarayoni bosqichma-bosqich bayon etilib, identifikatorlarni moslashtirish va yozuvlarni bog'lash, o'lchov birliklari va tasniflarni uyg'unlashtirish, panel ma'lumotlar shakllantirish, metadata boshqaruvi va versiyalar nazoratini o'z ichiga oladi. Shuningdek, empirik tahlilning shaffofligi va ishonchligini ta'minlash uchun zarur bo'lgan ma'lumotlar sifati tekshiruvlari, ya'ni yetishmayotgan qiymatlar, chet qiymatlar, ichki nomuvofliqlar, tanlov va bog'lash xatoliklari tahlil qilinadi. Natijalar bo'limida ma'lumotlar manbalari matritsasi, integratsiya xaritasi, proksi-indikatorlar katalogi va ma'lumotlar sifati kartasi kabi jadval va konseptual sxemalar keltiriladi. Maqola yakunida amaliy iqtisodiy tadqiqotlar uchun ishonchli, audit qilinadigan va qayta ishlab bo'ladigan empirik bazalarni yaratish bo'yicha aniq tavsiyalar beriladi.

Kalit so'zlar: empirik ma'lumotlar bazasi; ma'lumotlar manbalari; rasmiy statistika; korxonalar darajasidagi ma'lumotlar; so'rovnomalari; ma'lumotlarni integratsiyalash; yozuvlarni bog'lash; proksi-indikatorlar; ma'lumotlar sifati; metadata.

Аннотация: В статье разработана методология формирования эмпирической исследовательской базы данных на основе систематического отбора и интеграции различных источников информации, включая официальную статистику, данные на уровне предприятий (бухгалтерские, управленческие, транзакционные и реестровые записи), результаты опросов и прокси-индикаторы для трудноизмеримых экономических категорий. Предлагается практический набор критериев отбора источников данных, охватывающий репрезентативность, точность, охват, временную и пространственную сопоставимость, затраты на подготовку данных, правовые и этические ограничения, а также воспроизводимость результатов. Процесс интеграции данных описывается поэтапно и включает стандартизацию идентификаторов и связывание записей, гармонизацию единиц измерения и классификаций, формирование панельных данных, управление метаданными и контроль версий. Кроме того, обобщаются ключевые процедуры контроля качества данных, такие как анализ пропусков, выявление выбросов, проверка логической согласованности, а также оценка смещений отбора и ошибок связывания. В разделе «Результаты» представлены таблицы и концептуальные схемы, включая матрицу источников данных, карту интеграции, каталог прокси-индикаторов и шкалу оценки качества данных. В заключение сформулированы практические рекомендации по созданию надёжных, проверяемых и воспроизводимых эмпирических баз данных для прикладных экономических исследований.

Ключевые слова: эмпирическая база данных; источники данных; официальная статистика; данные предприятий; опросы; интеграция данных; связывание записей; прокси-переменные; качество данных; метаданные.

INTRODUCTION

The outcome of empirical economic research depends to a large extent on the question of which model was chosen, and even earlier—how the dataset was constructed—because econometric evaluations, even when methodologically sound, may suffer a substantial reduction in the reliability of their conclusions if they rely on data that are mismeasured, inconsistent, or incorrectly integrated. Therefore, the formation of the empirical base is not merely a “technical task” of research design, but a central component that determines the internal validity and external generalizability of scientific results [Wooldridge, 2010, 23–35].

In recent years, the palette of data sources used in economic analysis has expanded considerably. Official statistical data (national accounts, labour market indicators, price indices), enterprise- and bank-level transactions, tax and customs registers, electronic platforms, remote sensing data (geospatial information), questionnaires, and various proxy indicators (for example, night-time light intensity as a proxy for economic activity) are now widely employed. However, as the number of data sources increases, the complexity of empirical work also intensifies. Researchers face heterogeneity in units of observation (individual–enterprise–region), temporal consistency issues (monthly–quarterly–annual data), differences in classifiers (industry, region, product), the absence of common identifiers, confidentiality constraints, selection problems, and missing-data and linkage errors. As a result, “more data” does not necessarily imply “better data” [Cameron and Trivedi, 2005, 55–60].

The purpose of this article is to provide methodological foundations for the formation of an empirical research base, including:

1. criteria for selecting information sources;
2. the advantages and limitations of official statistics, enterprise data, questionnaires, and proxy indicators;
3. a practical algorithm for the data integration process;
4. quality control and documentation, including metadata management and reproducibility;
5. the construction of a “model empirical base” using illustrative tables and figures.

LITERATURE REVIEW

The methodological approach to constructing the empirical base follows the logic of “source → measurement → integration → verification” and is organized as a four-layer framework. The first layer, source selection, addresses the question of which data are relevant to the research problem, whether they provide sufficient scope and quality, and whether legal or institutional restrictions apply to their use. The second layer, measurement, focuses on the choice of indicators or proxies used to represent abstract concepts, such as “production efficiency.” Measurement error at this stage can introduce systematic bias and distort econometric estimates, as emphasized by Gujarati and Porter (2009, 98–105). The third layer, integration, concerns the combination, matching, or panel construction of data from different sources; this process itself can generate bias, commonly referred to as linkage bias (Imbens and Rubin, 2015, 74–80). The fourth layer, verification, involves assessing missing values, detecting outliers, performing cross-source comparisons, validation, and robustness checks. From an econometric perspective, data quality problems represent an “invisible” risk comparable to endogeneity. For example, classical measurement error typically attenuates estimated coefficients toward zero,

while non-classical error or sample selection may even reverse coefficient signs (Wooldridge, 2010, 299–305). Consequently, the empirical base constitutes the foundation of the identification strategy.

Official statistics produced by national statistical offices, ministries, and central banks are generally characterized by broad coverage, standardized methodologies, and continuity over time. At the same time, their high level of aggregation makes it difficult to identify micro-level mechanisms, while periodic updates of classifications and methodologies complicate intertemporal comparisons (UN, 2014, 11–15). In international practice, the core quality principles of official statistics—accuracy, timeliness, and consistency—are treated as key benchmarks (Eurostat, 2019, 7–10). In empirical research, official statistics are therefore best used as a source of background information and calibration, allowing researchers to verify macroeconomic trends, apply appropriate deflators, and standardize sectoral or territorial classifications.

Company-level information includes balance sheets, profit and loss statements, production logs, human resource systems, sales transactions, logistics records, and CRM data. Such data offer high accuracy and timeliness at the micro level; however, they are often affected by selection issues, limited representativeness, and low standardization. As business processes evolve, the definitions of indicators may also change over time, complicating longitudinal analysis (Cameron and Trivedi, 2005, 61–66). The availability of unique identifiers is critical for data integration: when VAT numbers or registration codes exist, linkage is straightforward, whereas their absence requires fuzzy matching based on attributes such as firm names, addresses, or telephone numbers. This approach increases the risk of linkage error and may distort the underlying sample structure (Angrist and Pischke, 2009, 46–52).

Questionnaire and survey data, including household and enterprise surveys, are particularly valuable for measuring unobservable characteristics such as motivations, expectations, or informal economic activity. Nevertheless, survey-based data are subject to a range of well-documented problems, including respondent errors, social desirability bias, recall bias, nonresponse, and issues related to sampling design (Groves et al., 2009, 71–90). When survey data are linked to official statistics, compatibility of units, weighting schemes, and aggregation levels becomes crucial to ensure internal consistency (Deaton, 1997, 32–40).

Proxy indicators are indirect measures used to represent concepts that cannot be observed directly. The selection of an appropriate proxy requires careful consideration of three key questions: how closely the proxy is related to the underlying construct in terms of validity; whether the proxy captures other confounding factors; and whether the proxy measure is stable across time and space (Wooldridge, 2010, 312–318). Proxies are often constructed in the form of composite indices. In such cases, international best practice recommends explicit conceptual modeling, transparent normalization and weighting procedures, and sensitivity analysis to assess robustness (OECD, 2008, 25–40).

The selection of information sources is guided by an eight-criteria framework that includes relevance to the research question, coverage in terms of units and time or geographic scope, precision and susceptibility to measurement error, comparability across classifiers, currencies, deflators, and standard units, the degree of granularity at the micro or macro level, the speed of updates and revision policies, legal and ethical compliance with confidentiality and consent requirements, and the capacity for replication through documentation, version control, and code availability. In practice, these criteria involve unavoidable trade-offs: for example, enterprise transaction data may exhibit high accuracy but limited representativeness, while official statistics typically provide broad coverage at the expense of micro-level detail.

Data integration relies on two principal approaches: deterministic linkage and probabilistic or fuzzy linkage. Deterministic linkage establishes one-to-one matches through unique identifiers such as register IDs, whereas probabilistic linkage relies on matching probabilities derived from names, addresses, phone numbers, and similar attributes. The latter involves an inherent trade-off between false matches and false non-matches (Groves et al., 2009, 210–220). Harmonization complements linkage by standardizing units of measurement, currencies, deflators, time aggregation (for example, monthly to quarterly), territorial codes, and industry classifications based on systems such as NACE or ISIC analogues (UN, 2014, 21–24).

REASERCH METHODOLOGY

The methodology of this study is based on a systematic approach to selecting, integrating, and analyzing heterogeneous data sources in order to form a reliable empirical base. Data are obtained from official statistical sources, enterprise-level administrative registers, financial statements, structured questionnaires, and proxy indicators derived from auxiliary datasets. Official statistics are used as the benchmark for definitions, deflators, and classifications, ensuring consistency and comparability across sources. Enterprise registers serve as the core identification framework, enabling deterministic linkage of firm-level financial and transactional data over time. Questionnaires are employed to capture unobservable characteristics, such as credit constraints and behavioral factors, with appropriate treatment of sampling design and nonresponse through weighting

adjustments. Proxy indicators are constructed only when direct measures are unavailable and are grounded in theoretical justification. Data integration involves harmonization of units, time alignment, and quality checks, including missingness analysis and outlier diagnostics. The empirical analysis applies panel-based econometric techniques, robustness checks, and sensitivity analysis to assess the stability and credibility of the estimated relationships.

ANALYSIS AND RESULTS

Many researchers view the linkage process as a purely technical step. In fact, if linkage errors are systematic, econometric conclusions become biased. For example, if entities that frequently change their name or move their address are more likely to “disappear” from the dataset, the resulting sample will be biased toward stable entities — a phenomenon known as survivorship bias [Angrist and Pischke, 2009, 73-77]. Therefore, linkage quality should be assessed using separate metrics, and groups with “successful” versus “unsuccessful” linkage should be systematically compared.

Microdata (enterprise-level data and survey questionnaires) often do not match official statistical aggregates. This mismatch does not necessarily imply that microdata are incorrect; rather, it may reflect differences in definitions, coverage, or time frames. Nevertheless, the researcher must clearly explain such discrepancies. For instance, microdata may cover only the formal sector, while official statistics may estimate aggregates using mixed or indirect methods. For benchmarking purposes, microdata are aggregated, and differences in both trends and levels are analyzed [Deaton, 1997, 44-50].

Surveys frequently rely on complex sampling designs involving stratification and clustering. If nonresponse rates increase over time or if the composition of respondents changes, simple OLS estimation may no longer yield unbiased results. Consequently, survey weights, design effects, and appropriate standard errors must be calculated and applied correctly [Groves et al., 2009, 320-340]. When constructing an empirical database, all sampling documentation should be preserved and stored as metadata alongside the data.

Proxy variables are often indispensable when direct measurements are unavailable. However, the risk of confounding increases when proxies are selected without strong theoretical justification. For example, internet penetration is frequently used as a proxy for innovation, yet it is also closely associated with income levels, urbanization, and education. Therefore, appropriate control variables and robust identification strategies are required [Wooldridge, 2010, 315-318]. Composite indices, in turn, may create the illusion of a single comprehensive measure while masking underlying sensitivities if robustness and sensitivity analyses are not conducted [OECD, 2008, 55-60].

When working with corporate and administrative data, it is essential to obtain informed consent, anonymize or pseudonymize data, use only the minimum necessary attributes, and ensure secure data storage. In particular, identifiers used for data linkage, such as names and addresses, are highly confidential. Researchers should therefore perform data integration within a secure environment and release analytical outputs containing only the minimum required information [UN, 2014, 32-35].

Below is a practical “constructor” for building an empirical database, illustrated using tables and figures. As an example, the topic of enterprise productivity and financial constraints is considered; however, the proposed structure is applicable to a wide range of empirical research topics (Table 1).

Table 1. Information sources matrix: task–source correspondence

Research component	Ideal information	Alternative source	Typical problem	Recommendation
Macro background (inflation, interest)	Official statistics / Central Bank	International Monetary Fund (IMF/ World Bank)	revision, difference in method	to «master» a resource
Enterprise output (Y)	Financial reporting + production	Tax return proxy	difference in definition	document the definition
Capital (K)	Fixed assets register	Balance (PPE)	reassessment	deflation, depreciation rule
Labor (L)	HR register	Questionnaire	informal workers	explain the difference in coverage
Financial constraint (X)	Credit history + limits	Loan application rejection (survey)	selection	triangulation of two sources
Innovation (M)	Patent/R&D information	IT Cost Proxy	confusion	index + sensitivity analysis
Area/network control	Classifiers	Address text	coding error	code dictionary + audit

This table briefly summarizes the key characteristics of data linkage across multiple sources. It presents the unit of analysis, key identifiers, binding types and directions, as well as linkage quality metrics for each source, thereby illustrating how different datasets are integrated and how the reliability of the linkage process is assessed in empirical research (Table 2).

Table 2. Integration map: “keys” and binding rules

Source	Unity	Key identifier	Binding type	Binding direction	Linkage quality metric
Official statistics	area–network–time	code (region, industry, t)	deterministic	aggregate → micro	compatibility (trend/level)
Enterprise register	enterprise	registry ID / TIN	deterministic	register → financial	match rate (%)
Financial report	enterprise–year	TIN + year	deterministic	finance ↔ register	duplicate check
Bank transaction	enterprise–month	TIN + month	deterministic	bank ↔ enterprise	coverage story
Questionnaire	enterprise	name/address/ phone	fuzzy	survey → register	precision/recall
Proxy (e.g. distance)	enterprise	geocode	deterministic	geodata → enterprise	geocode success

This overview summarizes commonly used proxy indicators in empirical analysis, outlining their calculation methods, main strengths, and inherent limitations. It also highlights the minimum set of controls required to mitigate biases such as reverse causality, sectoral heterogeneity, reporting distortions, and measurement errors, thereby supporting more robust and credible econometric inference (Table 3).

Table 3. Directory of proxy indicators

Construction	Proxy indicator	Calculation	Strength	Weak side	Minimum controls
Market competition	HHI/concentration	network shares	intuitive	definition sensitive	network FE, year FE
Financial constraint	Credit/development index	credit / asset	there is	reverse causality	lags, IV/DiD
Technological level	Electrical energy intensity	kWh / output	easy	sector difference	sector×year FE
Informality	Cash share	cash / sales	alarm	reporting bias	audit indicator
Productivity	TFP proxy	Cobb-Douglas residual	standard	measurement error	deflator, outlier trim

This overview outlines key data inspection procedures aimed at identifying selection risks, measurement errors, and integration problems. It emphasizes practical diagnostic tests, typical warning signals, and corrective actions, such as imputation, recoding, rule refinement, benchmarking, and version control, to ensure data consistency, transparency, and reliability in empirical analysis (Table 4).

Table 4. Data quality scorecard

Inspection	Purpose	Practical test	«Red flag»	Correction
Missingness profile	see selection risk	% missing by year/ sector	systematic missing	multiple imputation / restriction
Outlier diagnostics	error or extreme	percentiles, IQR	«jumps»	winsorize / audit
Consistency	definition compatibility	balance equations	asset≠liability	recoding
Linkage quality	match errors	match rate, manual sample	low match	rule update
Comparison	benchmarking	aggregate vs formal	big difference	annotation or calibration
Revision tracking	version control	save a snapshot	result change	versioning

This scheme presents a star-style data structure for empirical research, with a central fact layer capturing outcomes, treatments, and controls over firm–time units. It is linked to descriptive dimensions covering firms, time periods, regions, industries, and data sources, supporting consistent integration, traceability, and flexible panel econometric analysis(Figure 1).

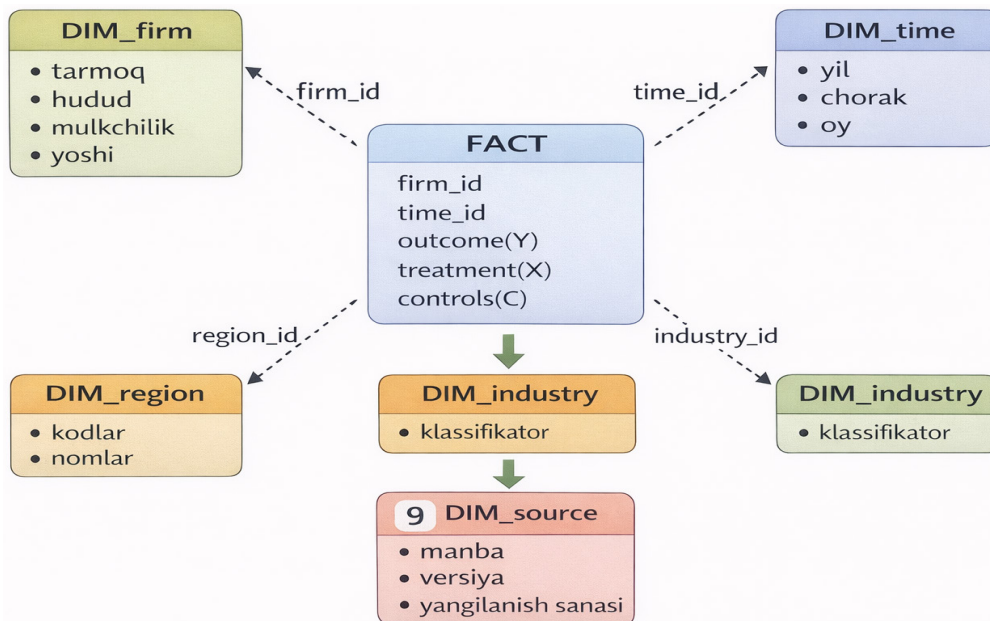


Figure 1. The concept of a “star schema” for data integration

This workflow outlines a stepwise record-linkage procedure that begins by checking the availability of unique identifiers. Deterministic matching is applied when IDs exist, while fuzzy matching is used otherwise, followed by threshold scoring, manual audits, duplicate checks, and rule adjustments. The process makes linkage errors explicit, measurable, and well documented for transparent empirical analysis(Figure 1).

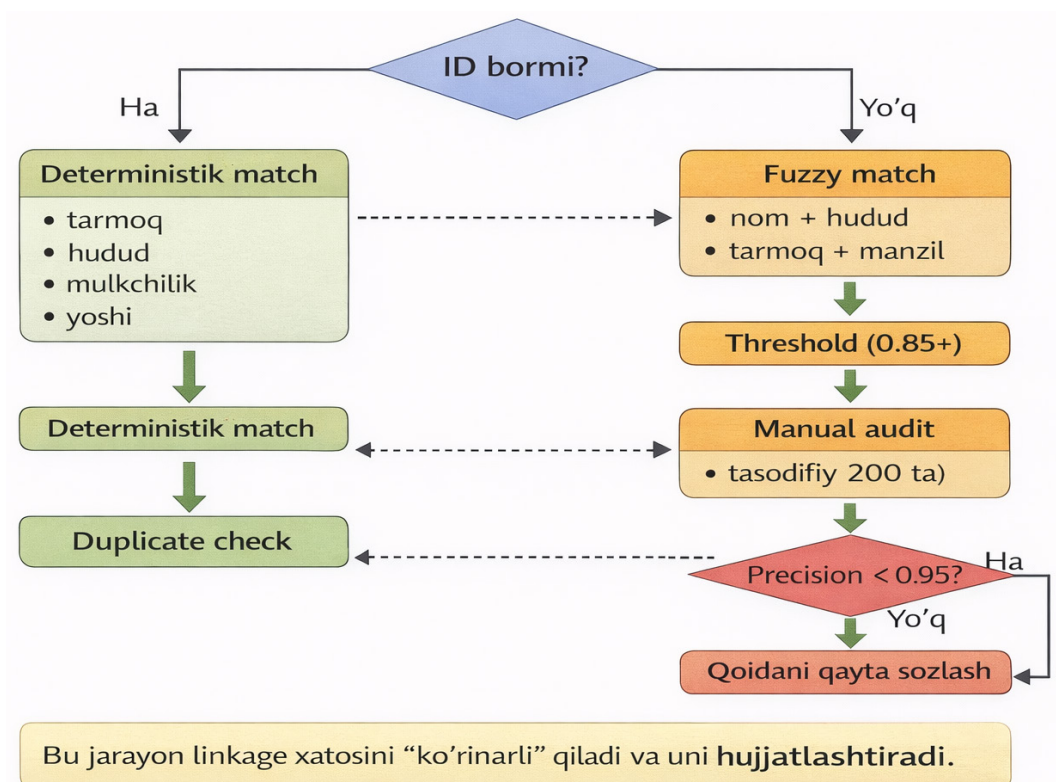


Figure 2. Linkage quality assessment (decision tree)

Final practical conclusion

When an empirical base is constructed in accordance with the above protocol, the following results are typically obtained:

- Deflators and classifiers are selected as “masters” through benchmarking with official statistics;
- the enterprise register becomes the “identifier core”, to which all data resources are linked;
- the “unobservable” component of financial constraints (for example, rejected credit applications) is incorporated through questionnaire data;
- proxy indicators are used as a “second signal” for mechanism identification or robustness checks;
- metadata management and versioning ensure the reproducibility and transparency of the results.

CONCLUSION AND SUGGESTIONS

The reliability of econometric results is ultimately determined by a set of selection and integration decisions made during the construction of the empirical base. This article leads to the following main conclusions.

First, the criteria for source selection must be explicit and transparent, including relevance, coverage, accuracy, comparability, granularity, timeliness, legal and ethical compliance, and replicability. Second, official statistics serve as the primary foundation for standardization and benchmarking; however, they do not fully capture underlying micro-level mechanisms. Third, enterprise-level data provide high micro-precision, yet strong differences in selection and definition remain, making enterprise registers and stable identifiers critically important.

Fourth, surveys are essential for measuring unobservable characteristics, but they require careful treatment of nonresponse, weighting, and complex sampling designs. Fifth, proxy indicators are informative only when grounded in theory and accompanied by sensitivity analysis; otherwise, the risk of confounding effects increases. Sixth, when integration steps — including key construction, harmonization, linkage, panel formation, quality control, and metadata management — are systematically documented, the resulting empirical base becomes auditable and repeatable.

As a practical recommendation, each study should include a dedicated “Data Appendix” describing data sources, measurement definitions, linkage rules, quality checks, and version histories. In addition, code and dataset versions should be stored separately, and a linkage audit should be conducted, even on a minimal sample. Under these conditions, the empirical base not only supports robust research results but also becomes a reusable scientific resource for future studies.

REFERENCES

1. Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.
2. Cameron, A. Colin, and Pravin K. Trivedi. 2005. *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press.
3. Deaton, Angus. 1997. *The Analysis of Household Surveys: A Microeconomic Approach to Development Policy*. Baltimore: Johns Hopkins University Press.
4. Eurostat. 2019. *European Statistics Code of Practice*. Luxembourg: Publications Office of the European Union.
5. Groves, Robert M., Floyd J. Fowler Jr., Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2009. *Survey Methodology*. 2nd ed. Hoboken, NJ: Wiley.
6. Gujarati, Damodar N., and Dawn C. Porter. 2009. *Basic Econometrics*. 5th ed. New York: McGraw-Hill.
7. Imbens, Guido W., and Donald B. Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge: Cambridge University Press.
8. OECD. 2008. *Handbook on Constructing Composite Indicators: Methodology and User Guide*. Paris: OECD.
9. United Nations. 2014. *Fundamental Principles of Official Statistics*. New York: United Nations.
10. Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. 2nd ed. Cambridge, MA: MIT Press.

Proofreader: Zokir ALIBEKOV

Layout and Designer: Oloviddin Sobir ugli

2026. № 1

© When materials are reproduced, the INNOVATION SCIENCE AND TECHNOLOGY journal must be cited as the source. Authors are responsible for the accuracy of the information in materials and advertisements published in the journal. Editorial opinions may not always align with those of the authors. Submitted materials will not be returned to the editorial office.

To publish articles in this journal, you may submit articles, advertisements, stories, and other creative materials through the following links. Materials and advertisements are published on a paid basis.

You may subscribe to the journal at any time using the following details. Once subscribed, please send a screenshot or photo of your payment confirmation to our Telegram page @iqtisodiyot_77. Based on this, we will send the latest issue of the journal to your address each month.

“The journal “INNOVATION SCIENCE AND TECHNOLOGY” has been registered by the Agency for Information and Mass Communications under the Administration of the President of the Republic of Uzbekistan from 09.10.2024 under the registration number №390637. License number: C-5669633. PNFL: 30407832680027

Our address: Tashkent city, Yunusobod district, 19th block,
House 17.



Acceptance of articles
Published every
monthly



Directions
Social, economic, political,
technological, scientific

 **Scopus || Scientific electronic journal specializing in Scopus**

CERTIFICATE NUMBER: №390637

**ORDER NUMBER ACCORDING TO
THE LICENSE REGISTER: C-5669633**

CONTACT:

 Contact us
+998 50 737 87 88

 Telegram channel
t.me/scopus_IST2100

 Journal official website
<https://ist-journal.uz/index.php/IST>