Integration into the Industrial Chain and Enterprise Innovation: A Novel Approach of Industrial Chain Measurement Using Firm Data

Lin Chen, Zhen Chen, Xiwen Zhang, Fei Long*

"The deep integration of the innovation chain, industry chain, capital chain, and talent chain" is a major national strategy at present, and how to achieve deep integration among these different chains has important theoretical and practical value. However, empirical frameworks and measurement methods for quantitatively analyzing the industrial chain are still lacking. Building on the approach of Acemoglu *et al*. (2012), this study uses artificial intelligence to gather related transaction and equity investment data from AiQiCha, ShangShangCha, regional tendering and bidding public service platforms, and the annual reports of listed enterprises, and then employs a depth-first search algorithm to analyze the industrial chain positions of enterprises. On this basis, combining data from input-output tables and the theory of directed graphs, this study calculates the industrial chain linkage, and further integrates the industrial chain centroid degree to measure the degree of industrial chain integration. The empirical results demonstrate that enterprises' integration into the industrial chain significantly improves their innovation performance via knowledge spillover and scale effects, and this improvement exhibits significant heterogeneity due to enterprise characteristics. Our analytical process and measurement results offer a relatively scientific quantitative analysis model to construct a modern industrial system and advance industrial chain development.

Keywords: industrial chain, enterprise innovation, industrial chain linkage, integration of innovation and industrial chains

1. Introduction

The report of the 20th National Congress of the Communist Party of China

* Lin Chen, Research Institute of Industrial Economy, Jinan University, Modern Industrial New System Research Base, School of Public Administration, Email: charlielinchen@qq.com. Zhen Chen, Science and Technology Strategy Consulting Research Institute of the Chinese Academy of Sciences, School of Public Policy and Management, University of Chinese Academy of Sciences, Email: zhenchen06@126.com. Xiwen Zhang, Research Institute of Industrial Economy, Jinan University, Email: zhangxiwen0218@qq.com. Fei Long, Research Institute of Industrial Economy, Jinan University, Email: longfeilydia@foxmail.com. This paper is supported by the Major Research Project of the Ministry of Education in Philosophy and Social Sciences (24JZD042, 22JZD007), National Natural Science Foundation Project (72473053, 72303045), and National Social Science Fund Project (23CJL033). Thanks to the Supercomputing Cloud Service of the Chinese Academy of Sciences Supercomputing Center, the author assumes full responsibility for the content.

emphasizes the need to "promote the deep integration of the innovation chain, industry chain, capital chain, and talent chain." *The Outline of the 14th Five-Year Plan (2021–2025) for Economic and Social Development and Long-Range Objectives through the Year 2035* of the People's Republic of China also indicates the need to "promote innovation among various enterprises across the upstream and downstream of the industry chain" and "support enterprises to integrate into the global industrial and supply chain." China proposes to integrate the individual innovation behaviors of enterprises into the industrial chain and the innovation chain from the macro perspective of "four chains integration" and the micro perspective of enterprise integration. This combination of macro and micro perspectives in the top-level design of the industrial chain is the only way to build a modern industrial system and implement an innovation-driven development strategy.

Balancing macro and micro perspectives, this paper focuses on the following core issues. At the macro level, it addresses how to construct an indicator system and a measurement framework for China's entire industrial chain based on micro-level enterprise data. At the micro level, it considers whether the domestic industrial chain in which specific enterprises are embedded, as well as their degree of integration, has an impact on enterprise innovation behavior, thereby driving innovation-driven development in the national economy from the bottom up and from individual enterprises to the entire supply side. Current domestic research on industrial chains, while focused mostly on theoretical discussions (Ren and Hong, 2005), has relatively few quantitative empirical investigations. Nevertheless, pioneering explorations have been conducted by, Zhang *et al*. (2020). Overall, there remains a relative lack of theoretical analysis for constructing industrial chain measurement models and comprehensive quantitative research on entire industries and full industrial chains based on micro-level enterprise data.

To this end, this paper attempts to make marginal contributions from the following dimensions. First, it approaches industrial chain research from a quantifiable perspective by constructing a theoretical framework and a specific methodology for industrial chain measurement. The input-output relationships among enterprises within an industrial chain—that is, the transformation of raw materials into final products—constitute a key challenge in quantitative industrial chain research. Neither input-output table theory at the industrial level nor supply chain management studies focusing on an individual or a limited number of enterprises can precisely characterize indicators such as industrial chain integration. On the one hand, while input-output table theory can depict interindustry input-output relationships at the macroeconomic level, its highly aggregated nature makes it difficult to reflect the degree of interconnectedness among micro-level enterprises. On the other hand, although supply chain research has examined the upstream and downstream linkages of individual and multiple enterprises, it has struggled to capture the characteristics of the entire industrial chain. Building on the approach of Acemoglu *et al*. (2012), which employed directed graph theory from mathematical sciences to analyze the construction of input-output tables, and drawing upon the methodologies of Lemelin (1982) and Fan and Lang (2000) for assessing intra-enterprise industrial chain linkages within business groups, this study characterizes economy-wide industrial chain connections based on Chinese enterprise-level data. This provides a relatively scientific theoretical framework and measurement methodology for industrial chain research.

Second, this study constructs a comprehensive industrial chain map based on micro-level enterprise data. Utilizing artificial intelligence technologies to extract data from enterprise information platforms (e.g., Qichacha, Shangshangcha), regional bidding announcements, and listed companies' annual reports, we obtained approximately 43.392 million enterprises with established relationships. By incorporating enterprises, suppliers, clients, and other economic entities, as well as applying the Depth-First Search[1] algorithm from graph theory, we identified roughly 5.369 million industrial chains. The construction of this industrial chain map database required approximately 2,000 core hours of computing power. Simultaneously, by comprehensively considering each enterprise's position within the industrial chain, its relative importance, and the closeness of its connections, this study innovatively constructs two key indicators: the centroid degree of industrial chains and industrial chain embeddedness. This represents a groundbreaking attempt to investigate industrial chains using domestic micro-level enterprise data and addresses the limitations of existing input-output tables in precisely characterizing dynamic micro-level interenterprise relationships. Moreover, it introduces a novel micro-econometric perspective to industrial chain research: a complex dynamic network system in which production nodes are intricately connected through material-flow-based industrial linkages, forming multilayered nested input-output relationships and ownership structures (property rights relationships).

Third, this study empirically verifies the technological innovation effects of enterprises' embeddedness in local industrial chains, with conclusions demonstrating how local industrial chain integration enhances enterprise innovation performance. This research also presents an investigation from the perspective of innovation chain and industrial chain convergence. While the "four chains integration" of innovation chains, industrial chains, capital chains, and talent chains has long remained at the theoretical and qualitative analysis stage, this paper conducts preliminary exploratory work in this domain.

---

[1] The depth-first search algorithm is a graph traversal algorithm in graph theory that systematically explores all branches of a spanning tree. Its fundamental principle involves starting from an initial node, proceeding as far as possible along one path until no new nodes can be accessed, then backtracking to explore alternative untraversed paths until complete graph traversal is achieved.

2. Literature and Hypotheses

2.1. Relevant Literature Review

In recent years, the innovation effects of industrial chains have emerged as a key research focus in both industrial organization theory and innovation theory. In early research on the industrial chain, more attention was paid to logistics and capital flow throughout the product chain, from natural resources to intermediate products to final products. As research has deepened, some studies have identified the existence of so-called technology-dominated industrial chains (Ren and Hong, 2005), proposing that industrial chains essentially represent techno-economic linkages formed among various industrial sectors along the path from raw material production to final consumption. From the perspective of innovation theory, since Schumpeter, academic research on the determinants of enterprise innovation has predominantly emphasized the importance of internal production conditions, such as capital, labor, and institutions, while relatively less attention has been paid to external factors, such as the industrial chains in which enterprises are embedded and their role in innovation output.

The current literature primarily examines the relationship between industrial chains and innovation activities through the following dimensions. First is the enterprise's position within domestic industrial chains. It is generally believed that enterprises leveraging their technological innovation and resource integration capacities can establish long-term strategic alliances with other enterprises in the industrial chain (Ren and Hong, 2005). Enterprises embedded in such networks benefit doubly: They can enhance their technological innovation capabilities by absorbing applied knowledge of cutting-edge technologies through production practice after obtaining high-quality intermediate inputs from upstream suppliers (Fan *et al*., 2023; Lai and Li, 2023), while also being driven to innovate through the reverse forcing effect of downstream enterprises' innovations (Wang *et al*., 2010; Dai *et al*., 2024). Furthermore, Zhang *et al*. (2020) quantified enterprises' positions within industrial chains and found that upstream enterprises tend to pursue radical innovation whereas downstream enterprises lean toward incremental innovation. Second, existing studies have examined how participation in global vertical specialization affects enterprises' innovation levels (Chen and Zhu, 2008). These studies revealed that international industrial chain linkages primarily influence innovation activities in industries with distant technology gaps and enterprises with medium to high human capital levels but that domestic linkages play a more significant role in shaping innovation in industries with proximate technology distances.

Moreover, with the increasing refinement of the industrial division of labor, enterprises gradually exhibit highly developed network characteristics (Chu *et al*., 2023). In the real world, industrial chains are no longer simple direct upstream-downstream relationships but rather constitute a “complex dynamic network system” (Diem *et al*., 2024) encompassing enterprises, suppliers, customers, and other economic entities and stakeholders. Accordingly, a third strand of the literature approaches industrial chain research from the perspective of production networks, extending the analysis to indirect interenterprise connections and preliminarily examining how production network embeddedness affects enterprise innovation (Wang and Hu, 2020). For instance, Bellamy *et al*. (2014) found that high network accessibility and interconnectivity have significant positive effects on enterprises’ technological innovation, with network interconnectivity further amplifying the positive impact of accessibility on technological innovation. Based on an analysis of 1,048 Chinese listed companies, Wang *et al*. (2023) discovered that supply chain network power and network cohesion significantly drive enterprise green technology innovation. Domestically, building upon the work of Acemoglu and Azar (2020) and Bigio and La’O (2020), Liu (2022) incorporated innovation into the production network framework, revealing that changes in production input structure can stimulate enterprise innovation through the productivity transformation of innovation and product price fluctuations.

In summary, existing research has primarily investigated the innovation effects of industrial chains and production networks across three dimensions—enterprises’ positions in domestic industrial chains, their participation in global vertical specialization, and complex production networks—providing an important foundation for subsequent academic research. However, the following aspects warrant further research. Regarding the research subject, industrial chains should be regarded as complex networks formed by the interweaving of economic entities and stakeholders with multilayered nested relationships. With the exception of the third strand of the literature, which preliminarily examines the innovation effects of production networks, most studies have focused solely on direct upstream–downstream relationships, thereby underestimating the “cascade (amplification) effects” of technological innovation arising from higher-order industrial linkages in production networks (Acemoglu *et al*., 2012; Acemoglu and Azar, 2020; Liu, 2022). Recognizing that shareholders and other stakeholders influence enterprise production and operational decisions, this study expands upon the third research strand of literature by incorporating interenterprise investment and financing relationships into the production network, thereby constructing a multilayered industrial chain network that encompasses both transactional and financial linkages. From a research perspective, this study focuses on the innovation effects of enterprises’ embeddedness in regional industrial chains. Existing research has predominantly examined enterprise innovation activities and performance through national or global supply chains and production networks, neglecting the critical influence of geographical proximity between transacting parties on innovation knowledge transfer. This oversight may lead to misjudgments regarding the innovation effects of industrial chain integration. Accordingly, following established research trajectories while addressing this gap, this paper specifically investigates how local industrial chain integration affects technological innovation, providing a more precise evaluation of industrial chains' innovation effects through a regional industrial chain perspective.

2.2. Theory and Hypotheses

2.2.1. Enterprise Integration into the Industrial Chain and Technological Innovation

The concept of industrial chains can be traced back to Adam Smith's theory of labor division, which, from his 1776 work *The Wealth of Nations*, posits that labor specialization through the division of labor can significantly enhance production efficiency. Subsequently, Porter's *Competitive Advantage* built upon this theory to develop the value chain concept, emphasizing how enterprises achieve competitive advantages through a series of value-adding activities. Correspondingly, the industrial chain is defined as the chain formed by the various stages of a specific industry, from the acquisition of raw materials, production and processing, and sales to final consumption, following the sequence of product production and adding value. Each segment consists of multiple enterprises interconnected through input-output relationships, collectively accomplishing product value creation and accumulation. Consequently, industrial chain integration provides enterprises with crucial pathways for exploring potential markets and establishing mutually beneficial partnerships with collaborators. By consolidating diverse resources, enterprises can effectively enhance their overall innovation performance through such integration. Specifically, on the one hand, the tightly connected network within industrial chains enhances enterprises' ability to identify external opportunities, helping them discern the strengths and weaknesses of potential partners and evaluate the feasibility of different innovation pathways. This not only reduces search costs and trial-and-error expenses in collaborative R&D activities but also effectively mitigates innovation risks. Moreover, the shared objectives among enterprises in the chain help diminish opportunistic behavior among partners and increase mutual benefits. On the other hand, enterprises can acquire substantial valuable knowledge and resources from the industrial chain to build their innovation resource pools. By accessing broader market and user demand information, they can clarify innovation directions and improve their knowledge absorption and transformation capabilities. Based on this, we propose the following hypothesis:

Hypothesis 1: The deeper the enterprise's integration within industrial chains, the more conducive it is to enhancing innovation performance.

2.2.2. The Knowledge Spillover Effect of Enterprises Integrating into the Industrial Chain

Innovation economics theory posits that innovation activities depend on enterprises' effective acquisition of internal and external knowledge. As technology and knowledge grow increasingly complex, enterprises relying solely on internal knowledge absorption and integration struggle to meet modern innovation demands and adapt to rapid market changes (Qian *et al*., 2010). Consequently, enhancing external knowledge acquisition becomes a primary source for maintaining competitive innovation advantages. Industrial chains serve as crucial channels for obtaining external information and knowledge as well as important platforms for interenterprise collaborative R&D (Chu *et al*., 2019). They effectively help enterprises acquire new knowledge and fill technological gaps (Andersson *et al*., 2005) and demonstrate significant knowledge spillover effects. Benefiting from these spillovers, enterprises embedded in industrial chains should theoretically exhibit superior external knowledge acquisition capabilities compared to isolated enterprises. Specifically, prior researches define an enterprise's ability to identify, acquire, comprehend, and apply external knowledge as absorptive capacity, which encompasses three critical stages: recognition, assimilation, and application (Zahra and George, 2002). During the recognition stage, an enterprise's ability to identify external knowledge functions as an extension of its prior related knowledge. Enterprises within industrial chains, due to their production technology linkages, possess an inherent understanding of other members' technologies and knowledge, thereby exhibiting stronger capabilities in identifying valuable information. In the assimilation stage, enterprises must systematically process, interpret, and comprehend external information. Since certain external knowledge requires specific technological environments for implementation, enterprises that lack dedicated assets for knowledge absorption face significant challenges in understanding and replicating such external knowledge. Enterprises that operate within the same industrial chain inherently share operational contexts and production technology linkages, which helps them overcome various technology–incompatibility barriers, and they consequently demonstrate stronger capabilities in assimilating external knowledge. During the application stage, enterprises must monitor market dynamics and external environmental changes to analyze and predict market responses to new products or services. Enterprises embedded in industrial chains are better positioned to observe market fluctuations and emerging technological information across upstream and downstream sectors. Consequently, their innovation direction and content can more accurately align with industrial development needs, facilitating the introduction of technologies and products that conform to future trends. Based on this analysis, we propose the following hypothesis:

Hypothesis 2: Through knowledge spillover effects, deeper integration within industrial chains effectively facilitates interenterprise innovation knowledge sharing, thereby significantly enhancing enterprise innovation performance.

2.2.3. The Scale Effect of Enterprises Integrating into the Industrial Chain

The Schumpeter hypothesis suggests that larger enterprise size drives technological progress. However, modern enterprise development evolves beyond internal growth alone, increasingly expanding through business interactions, investment channels, and comprehensive industrial chain integration both upstream and downstream. Through such vertical cooperation within industrial chains, enterprises can leverage partners' distribution networks and customer bases to identify new market opportunities and expand their business reach. This approach not only enhances market penetration for existing products but also provides valuable market feedback and consumer insights, creating opportunities to access emerging markets. It follows logically that deeper industrial chain integration facilitates market expansion for enterprises, thereby incentivizing production scale increases that achieve economies of scale and free up additional resources for R&D investment. Due to high-order connectivity within industrial chains (Acemoglu *et al*., 2012), when one enterprise expands production, it generates increased demand for upstream suppliers' products, creating a multiplier effect that propagates economies of scale throughout the chain. Furthermore, market expansion implies greater potential scale for innovative products, which elevates enterprises' expected innovation returns while shortening investment payback periods and ultimately encourages greater capital allocation to high-risk, high-reward innovation activities. Based on this mechanism, we propose the following hypothesis:

Hypothesis 3: Through scale effects, deeper integration within industrial chains effectively stimulates enterprise innovation activities, thereby significantly enhancing enterprises' innovation performance.

3. Theoretical Framework, Indicator System, and Data Processing for Measuring the Industrial Chain

3.1. Data Preparation: Constructing a Whole-Industry-Chain Network Map Based on the Depth-First Search Algorithm

First, we utilized Python programming to access the supercomputing center of the Computer Network Information Center of the Chinese Academy of Sciences and extracted enterprise relationship data from open-source platforms including Qichacha and Shangshangcha. In this process, we collected the publicly available enterprise maps, relationship networks, equity structures, and annual reports of listed companies from the

CSMAR database. The extracted data encompassed suppliers, clients, external investments, holding companies, and shareholders, from which we established the foundational "nodes" for constructing our subsequent "industrial chain mapping database."

Next, partially adopting the high-order industrial linkage approach, we utilized the depth-first search algorithm from graph theory to construct quantifiable industrial chains. As previously mentioned, the collection of nodes and edges followed the high-order industrial linkage framework, specifically implementing m+1 order linkages. Specifically, this means that starting from a given node, we established an undirected edge that connected to subsequent nodes whenever relationships were identified in the industrial chain mapping database. The algorithm then iteratively performed second-order searches, identifying all related nodes and edges for each subsequent node until all m+1 order connections originating from the initial enterprise were exhaustively mapped, thereby constructing the first complete industrial chain. The algorithm then iteratively proceeded to select enterprises outside existing chains and exhaustively mapped their infinite-order industrial linkages to construct subsequent chains. This process continued recursively until all nodes were incorporated into independent industrial chains. For measurement purposes, our analysis confined these industrial chains within prefecture-level city boundaries, ultimately establishing 5,368,908 (approximately 5.369 million) chains across 315 prefecture-level and higher cities, using approximately 2,000 core hours of computational resources.

To facilitate subsequent industrial chain indicator measurement, we first compiled a baseline list of industrial enterprises (1998–2014) with their enterprise entity codes. We then obtained unified social credit codes for the above-scale enterprises from China's business registration database. After merging with the China Industrial Enterprise Database and applying Chen (2018)'s methodological framework, we matched our proprietary industrial chain mapping database with listed companies using unified social credit codes. This process yielded 371,620 relationship enterprises with complete financial and operational data and formed 111,946 industrial chains for indicator calculations.

3.2. Measurement Framework and Indicator System

3.2.1. The First-Order Industry Linkage Measurement of the Relevance of Enterprise Industry Chains (*ind_link*)

This study drew on the methodological insights of Fan and Lang (2000) and Acemoglu *et al*. (2012) to measure industrial chains based on first-order industrial linkages, utilizing *Input-Output Tables of China* as a foundation. China conducts nationwide input-output surveys every 5 years, in years ending with 2 or 7, compiling benchmark input-output tables across sectors. Given our data coverage from 1998 to

2014, we utilized the simplified *2007 National Input-Output Table with 135 Sectors* for analysis.

$$A_1 = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix}_{n\times n} . \tag{1}$$

In Equation (1), A_1 represents the direct input coefficient matrix from traditional input-output tables, where N denotes the set of industries in the industrial chain, $N = \{1, 2, 3\ldots\ldots n\}$, with $i, j \in N$ and n being the total number of industries. Each element a_{ij} constitutes the direct input coefficient between industry i and industry j, indicating the quantity of industry i's products consumed in producing one unit of industry j's output.

First, we calculated the weighted industrial outdegree (a_i) for node (enterprise) i by summing all outward edges and their corresponding weights. This involved aggregating the direct consumption coefficients $a_{ij}\,(\sum_{j=1}^{n} a_{ij})$ between industry i and all other industries. The weighted industrial outdegree quantifies the extent to which industry i's products are directly consumed as intermediate inputs across various industries within the industrial chain. The collection of elements in column vector A_2 constitutes the degree sequence of economy (representing the edge set aggregation, which can be interpreted as measuring each industry's aggregate influence on the national economic industrial chain), yielding Equation (2).

Acemoglu *et al*. (2012) further added up the weighted industrial outdegrees to the averaged industrial chain connectivity degree $ind\_link^{/}$, where C_1 represents $1\times n$ row vector composed of elements $1/n$. This mean-value approach disregards interindustry scale differences, yielding Equation (3). Building upon this foundation, the coefficient of variation can be derived. When accounting for higher-order industrial linkages, this framework extends to obtain second-order and m+1-order industrial connectivity coefficients.

$$A_2^T = \begin{bmatrix} \sum_{j=1}^{n} a_{1j} & \sum_{j=1}^{n} a_{2j} & \cdots & \sum_{j=1}^{n} a_{nj} \end{bmatrix}_{1\times n} = \begin{bmatrix} a_1^n & a_2^n & \cdots & a_n^n \end{bmatrix}_{1\times n} , \tag{2}$$

$$ind\_link^{/} = C_1 \times A_2 . \tag{3}$$

Drawing on Fan and Lang (2000)'s methodology of calculating integration indicators using operating revenue as weights, this study further assigned coefficient c_i to each node, with the weight defined as "the ratio of industry's core business revenue to the total core business revenue across the chain." This yielded the following equation, where $\sum_{i=1}^{n} c_i = 1$:

$$C_2 = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \end{bmatrix}_{1\times n}, \tag{4}$$

$$ind\_link = C_2 \times A_2. \tag{5}$$

Based on the submatrix of direct input coefficients from the input-output table and following Equation (5), we calculated the industrial chain connectivity indicator (ind_link) for each chain. This indicator quantitatively measures the degree of interdependence among sectors or enterprises within an industrial chain. A higher ind_link value indicates that the sector's output participates more extensively in other sectors' production processes, reflecting stronger connectivity within the industrial chain. It should be noted that while some bidding samples include transaction amounts, most samples cannot determine the magnitude of material flows or raw material supply between enterprises. However, to maximize the sample size, this study did not limit the data to bid announcement records with explicit input-output directions, thereby resulting in undirected graph data. Ultimately, by integrating the material flow data from input-output tables with directed graph theory, we calculated the industrial chain connectivity indicator incorporating input-output directional information.

3.2.2. Measurement of the Centroid Degree of the Enterprise's Industrial Chain d_f Based on the Theorem of Centroid Motion

Building upon Antràs *et al*. (2012)'s industry upstreamness measure, this study constructed an enterprise-level centroid degree indicator to estimate an enterprise's position as the "centroid" (or core) within industrial chains. The upstreamness indicator quantifies a sector's relative position along the complete value chain (from raw materials to final products) by measuring the average distance between the sector's output and final products. Higher upstreamness values indicate a greater average distance from final consumption, meaning the sector's products undergo more production stages before reaching end consumers, while lower values suggest fewer intermediate stages.

First, we constructed the industry upstreamness indicator u_i to quantify sector *i*'s position within the entire industrial chain, as follows:

$$u_i = 1\times\frac{X_i}{Y_i} + 2\times\frac{\sum_{j=1}^{r} a_{ij}X_j}{Y_i} + 3\times\frac{\sum_{j=1}^{r}\sum_{k=1}^{r} a_{ik}a_{kj}X_j}{Y_i} + \ldots \tag{6}$$

Y_i represents the total output of sector *i*. X_i denotes the portion of sector *i*'s output allocated to final consumption. a_{ij} indicates the required input from sector *i* to produce one unit of sector *j*'s output. u_i is the weighted average of the proportion of total output that is served as final products and intermediate goods at various production stages within sector *i*. A higher u_i value indicates greater upstreamness, reflecting a longer distance between intermediate goods and final products and a broader influence on downstream sectors. Notably, $u_i \geq 1$, with equality holding if and only if all of sector *i*'s output serves as the final products.

Although the upstreamness indicator effectively characterizes relative positions across industries within a given industrial chain, due to the differences in the total number of stages across different industrial chains, this indicator cannot be used to directly compare the number of stages experienced in the process, where products produced by enterprises in the same industry along different industrial chains are used as intermediate goods in subsequent production stages until they form final products. This study posits that an enterprise *f*'s relative position across chains can be quantified as the ratio between (1) the distance from enterprise *f* to the chain's centroid degree and (2) the chain's total length, with the computational methods specified in the following equations:

$$u_c = \sum_{i=1}^{n} u_i \times c_i, \tag{7}$$

$$d_f = \left(1 - \frac{\left|u_f - u_c\right|}{u_{max} - u_{min}}\right) \times production_f. \tag{8}$$

d_f represents enterprise *f*'s industrial chain centroid degree, where $\left|u_f - u_c\right|$ measures the distance between enterprise *f* and the chain's centroid degree. Here, u_f denotes the upstreamness of enterprise *f*'s sector, while u_c is calculated as the product of the upstreamness (u_i) of various industries and the proportion (c_i) of corresponding industries in the main business income on the chain. $(u_{max} - u_{min})$ represents the total length of the industrial chain, determined by the distance between the industry with the highest upstreamness (u_{max}) and the industry with the lowest upstreamness (u_{min}) in the industrial chain. Using Equation (8), we obtain enterprise *f*'s industrial chain centroid degree (d_f), where $production_f$ represents enterprise *f* 's output share relative to total chain output. A higher d_f value ($0 < d_f < 1$) indicates a smaller relative distance (*t*) between enterprise *f* and the centroid of the industrial chain, reflecting greater centroid degree within the industrial chain.

3.2.3. Measurement of the Degree of Integration of Enterprises into the Industrial Chain

The industrial chain linkage degree (ind_link) is a shared variable among enterprises within the same industrial chain, reflecting the degree of interconnectedness among the enterprises that constitute the chain. However, different enterprises within the same industrial chain occupy distinct positions, leading to varying capacities to either benefit from the chain's spillover effects or drive its development.

For example, enterprise A, which produces keyboards, and enterprise B, which produces chips, both belong to an electronic computer industrial chain and share the same industrial chain linkage degree (ind_link). However, their relationships with this industrial chain represent two extremes—clearly, the chip manufacturer holds a more central position in the chain. It is more capable of generating or receiving spillover effects (i.e., the cascade amplification effect described by Acemoglu *et al*., 2012) and has a greater influence in leading and shaping the development direction of this industrial chain. If we analogize the electronic computer industrial chain as a train, the chip manufacturer would undoubtedly be the locomotive, while the keyboard producer would be a lightly loaded railway carriage. A higher ind_link indicates stronger coupling and traction between the train's carriages. Although both are parts of the same train, they play fundamentally different roles and may face distinct exogenous shocks—just as in a train collision, the leading locomotive, with its engine, would inevitably sustain the most severe damage. If we further analogize the electronic computer industrial chain as an ox, the chip manufacturer would undoubtedly represent the ring in the ox's nose, which is capable of steering the entire system with minimal intervention. In contrast, keyboard enterprises may belong to the area placed on the thickest part of the cow leather, unable to significantly alter the ox's direction of movement.

In summary, the enterprise's industrial chain integration level is jointly determined by its own industrial chain centroid degree (d_f) and the overall industrial chain linkage degree (ind_link), as expressed in the following Equation (9):

$$ind\_chain_f = d_f \times ind\_link. \tag{9}$$

As previously demonstrated, a higher industrial chain linkage degree (ind_link) indicates stronger interenterprise connections within the entire industrial chain, which is an internal relational indicator of the industrial chain. Conversely, the industrial chain centroid degree (d_f) reflects an enterprise's influence and driving capacity in local industrial chain development, representing an external relational indicator between the enterprise and the industrial chain. The multiplication of these two indicators yields a comprehensive micro-level indicator (ind_chain_f) that simultaneously evaluates both the overall linkage tightness of an industrial chain within a specific geographic scope and the positional changes of individual enterprises within the industrial chain. In this study, this indicator is employed to represent the degree of enterprise-level industrial chain integration.

4. Empirical Design and Analysis of the Results

4.1. Econometric Model

To examine the impact of industrial chain integration on enterprises' innovation performance, the baseline model of this study is specified as follows:

$$innovation_{ft} = \alpha_1 ind\_chain_{ft} + \alpha_2 X_{ft} + \varphi_f + v_t + \varepsilon_{ft}, \tag{10}$$

where *f* denotes individual enterprises, *t* represents the statistical year, *innovation* measures enterprise innovation performance using annual patent applications, and *ind_chain* indicates industrial chain integration degree. *X* represents the set of control variables. The regression incorporates enterprise fixed effects φ_f and year fixed effects v_t, with ε_{ft} denoting cluster-robust standard errors at the enterprise level.

4.2. Variable Descriptions

Enterprise innovation performance is measured by the natural logarithm of annual patent applications (*innovation*1). Following Li and Zheng (2016), we further constructed *innovation*2 as the natural logarithm of the sum of invention patent applications and utility model patent applications from China's patent database, representing innovation performance aimed at technological advancement and competitive advantage acquisition.

To mitigate the estimation bias caused by omitted variables, the control variables were selected with reference to Lv *et al*. (2023), including enterprise size (*size*), enterprise age (*age*), and capital intensity (*kl*). Following Yang *et al*. (2015) and Li and

Wang (2017), we incorporated cost-profit ratio (*profitr*), subsidy income (*subsidy*), and asset-liability ratio (*debt*). Drawing from Chen and Zhu (2011), we selected industry competition level (*HHI*). Due to space constraints, detailed descriptive statistics and the data processing procedures are not reported in the main text, but interested readers may request them from the authors.

4.3. Benchmark Regression Results

Table 1 presents the regression results of industrial chain integration on enterprise innovation performance. Overall, the coefficient of industrial chain integration shows significantly positive effects on innovation performance, and the results remain robust with the inclusion of additional control variables. This indicates that higher levels of industrial chain integration contribute to enhancing innovation performance at the enterprise level, thereby providing preliminary validation for Hypothesis 1, which posits that industrial chain integration facilitates the improvement of enterprises' innovation performance.

Table 1. Integration Degree of the Industrial Chain and Enterprise Innovation Performance

Variable	*innovation*1	*innovation*2	*innovation*1	*innovation*2
	(1)	(2)	(3)	(4)
ind_chain	0.0600***	0.0494***	0.0440***	0.0362***
	(0.0051)	(0.0048)	(0.0052)	(0.0049)
Control Variables	No	No	Yes	Yes
Enterprise fixed effects	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Sample size	1840719	1840719	1737114	1737114
Adjusted R^2	0.4242	0.4262	0.4324	0.4351

Note: ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. Values in parentheses are cluster-robust standard errors at the enterprise level. The same applies to the tables below.

4.4. The Mechanism of the Degree of Integration in the Industry Chain Promoting Enterprise Innovation

4.4.1. Knowledge Spillover Effect

Following the methods of Byun *et al*. (2021) and Wang *et al*. (2023), technological proximity ($techprox_{ft}$) was used as a proxy for knowledge spillovers. The regression results in columns (1)–(3) of Table 2 show that the degree of industrial chain integration has a significantly positive impact on enterprises' technological proximity, patent citations, and family patent citations. This indicates that deeper integration into the industrial chain enhances enterprises' ability to absorb external knowledge, facilitates access to advanced technologies and knowledge from upstream and downstream enterprises, and creates broader opportunities for innovation activities. These regression results support Hypothesis 2.

4.4.2. Scale Effect

Drawing on Zweimuller and Brunner (2005), mechanism variables were constructed for enterprise scale effects, specifically including output growth, profit growth, and return on assets. Output growth was calculated as the natural logarithm of the difference between the current period's gross industrial output value and that of the previous period. Similarly, profit growth was measured by the natural logarithm of the difference between the current total profits and those of the prior period. Return on assets was computed as the ratio of total profits to total assets. The results in columns (4)–(6) demonstrate that industrial chain integration significantly enhances output growth, profit growth, and return on assets. This suggests that deeper integration into the industrial chain strengthens enterprises' production scale, profitability, and asset returns, providing partial support for Hypothesis 3.

Table 2. Mechanism Test

Variable	Knowledge spillover effect			Scale effect		
	Technological proximity	Patent citations	Family patent citations	Output growth	Profit growth	Return on assets
	(1)	(2)	(3)	(4)	(5)	(6)
ind_chain	3.6728***	0.2442***	0.3103***	0.6273***	0.5009***	0.0938***
	(0.2834)	(0.0062)	(0.0071)	(0.0207)	(0.0242)	(0.0223)
Control variables	Yes	Yes	Yes	Yes	Yes	Yes
Enterprise fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Sample size	1793678	1793678	1793678	889353	762462	1793362
Adjusted R^2	0.5802	0.6393	0.4015	0.3187	0.4713	0.4925

4.5. Further Analysis

4.5.1. Heterogeneity Analysis

First, for the identification of lead enterprises in the industrial chain, this study followed Acemoglu *et al*. (2022) by defining enterprises in the top 10% of the Bonacich centroid degree within the same relational network as industrial chain "leaders." Second, regarding ownership structure, enterprises were categorized into state-owned and non-state-owned enterprises based on their property rights. Third, for industry competition intensity, drawing on Chen and Zhu (2011), competitive environment of each industrial chain was measured by calculating the median Herfindahl index of relevant industries within each city annually. The regression results indicate that industrial chain integration does not exert a significant impact on the innovation performance of lead enterprises, state-owned enterprises, or enterprises in low-competition industries. Instead, it has a significantly positive effect on the innovation performance of other enterprises in the industrial chain. Due to space constraints, this paper does not report the endogeneity and robustness tests, further regression analyses, or detailed discussions in the main text. Interested readers may contact the authors for these additional results.

4.5.2. Extensibility Analysis of the Integration of the Innovation Chain and the Industrial Chain

Considering that knowledge, technology, and other innovation factors create value along industrial chains through products or services, this study adopted the research approach of Lai and Li (2023) and constructed an undirected weighted collaborative innovation network along industrial chains based on copatenting activities between enterprises and other enterprises within the same industrial chain. Using degree centrality, eigenvector centrality, and weighted eigenvector centrality to approximate the integration of innovation chains and industrial chains among enterprises, the regression results show that industrial chain integration has a significantly positive effect on innovation degree centrality and eigenvector centrality. This suggests that promoting the convergence of innovation chains and industrial chains represents a crucial pathway through which industrial chains enhance enterprises' innovation performance.

5. Main Conclusions and Policy Implications

How do participation in industrial chain development and integration into local industrial chains incentivize enterprises' innovation performance? To address this question, this study utilized artificial intelligence-powered distributed web scraping to collect enterprise relevance data, including bidding information, related-party transactions, and equity investments, from platforms such as Qichacha, Shangshangcha, regional bid announcements, and listed companies' annual reports. Utilizing approximately 2,000 core hours of computing power and depth-first search algorithms, we reconstructed enterprises' embedded industrial chains and established China's first comprehensive industrial chain mapping database. By integrating material flow data from input-output tables with directed graph theory, we calculated intrachain connectivity and centroid degrees for domestic industrial chains, ultimately deriving enterprise-level industrial chain integration indicators to examine their impact on enterprise innovation performance. The main conclusions are as follows. First, industrial chain integration significantly enhances enterprises' innovation performance, with this effect primarily driven by knowledge spillovers and scale effects. Second, this enhancement exhibits significant heterogeneity depending on enterprise characteristics (e.g., whether they are lead enterprises or their ownership structure). Third, the convergence of innovation chains and industrial chains objectively exists to some extent, demonstrating that technological innovation activities can achieve deep integration with large-scale industrial production.

The policy implications are as follows. First, promoting regional industrial chain development holds significant importance for establishing long-term mechanisms under an innovation-driven development strategy. While enterprises are the key entities of innovation activities, they should not rely solely on "individual heroism." Governments at all levels should formulate innovation incentive policies based on industrial chains. For individual enterprises' innovation strategies, enterprises should opportunistically develop appropriate innovation strategies that account for their position in the industrial chain, enterprise size, ownership structure, and external competitive environment (market structure). Second, enterprises should be encouraged to continuously integrate into industrial chains, particularly by leveraging the scale economies of lead enterprises and large enterprises, to enhance the resilience and security of industrial and supply chains. Our empirical evidence suggests that promoting the lead enterprise system has a theoretical foundation. Mega-sized enterprises and large state-owned enterprises (such as central state-owned enterprises) serving as lead enterprises in both industrial and innovation chains should actively embed themselves in industrial chains by optimizing their production networks, organizational structures, and management mechanisms, thereby guiding or even leading the development direction of industrial chains. Third, encouraging collaborative R&D among enterprises and optimizing the market competition environment are crucial approaches to fully realizing the innovation effects of industrial chains. Compared with some developed countries, although lead enterprises such as Huawei, BYD, and CATL demonstrate outstanding independent innovation capabilities, whether China, as the world's largest manufacturing country, achieves “1 + 1 > 2” synergistic effects in overall innovation performance through its industrial chains remains a critical question. Furthermore, whether small and medium-sized enterprises can obtain technological spillovers from other enterprises in the chain while cooperating with lead enterprises in industrial chain development requires empirical verification. Our preliminary results show the existence of such scale economies and technology spillover effects from collaborative R&D.

References

Acemoglu, D., & Azar, P. D. (2020). Endogenous Production Networks. *Econometrica*, 88(1), 33–82.

Acemoglu, D., Akcigit, U., & Celik, M. A. (2022). Radical and Incremental Innovation: The Roles of Firms, Managers, and Innovators. *American Economic Journal: Macroeconomics*, 14(3), 199–249.

Acemoglu, D., Carvalho, V. M., Ozdaglar, A., & Tahbaz-Salehi, A. (2012). The Network Origins of Aggregate Fluctuations. *Econometrica*, 80(5), 1977–2016.

Andersson, U., Björkman, I., & Forsgren, M. (2005). Managing Subsidiary Knowledge Creation: The Effect of Control Mechanisms on Subsidiary Local Embeddedness. *International Business Review*, 14(5), 521–538.

Antràs, P., Chor, D., Fally, T., & Hillberry, R. (2012). Measuring the Upstreamness of Production and Trade Flows. *American Economic Review*, 102(3), 412–416.

Bellamy, M. A., Ghosh, S., & Hora, M. (2014). The Influence of Supply Network Structure on Firm Innovation. *Journal of Operations Management*, 32(6), 357–373.

Bigio, S., & La'o, J. (2020). Distortions in Production Networks. *Quarterly Journal of Economics*, 135(4), 2187–2253.

Byun, S. K., Oh, J. M., & Xia, H. (2021). Incremental vs. Breakthrough Innovation: The Role of Technology Spillovers. *Management Science*, 67(3), 1779–1802.

Chen, L. (2018). Re-exploring the Usage of China's Industrial Enterprise Database. *Economic Review(Jingji Pinglun)*, (6), 140–153.

Chen, L., & Zhu, W. (2008). Policy Effect of Rebating Export Tax and Subsidizing Innovation in Innovation Incentive. *Economic Research Journal(Jingji Yanjiu)*, (11), 74–87.

Chen, L., & Zhu, W. (2011). Innovation, Market Structure and Administrative Entry Barriers: Empirical Test of Schumpeter's Hypothesis Based on Chinese Industrial Enterprise Data. *China Economic Quarterly*, (1), 653–674.

Chu, D., Li, Y., & Zhang, T. (2023). Research on the Causes of VAT Burden Difference from the Perspective of Industrial Interconnection. *Economic Research Journal(Jingji Yanjiu)*,(7), 174–190.

Chu, Y., Tian, X., & Wang, W. (2019). Corporate Innovation Along the Supply Chain. *Management Science*, 65(6), 2445–2466.

Dai, J., Liu, G., Ye, Z., Mao, H. & Ye, Y. (2024). Downstream Competition and Upstream Innovation: Theory and Evidence from China. *China Finance and Economic Review*, 13(1), 67-87.

Diem, C., Borsos, A., Reisch, T., Kertész, J., & Thurner, S. (2024). Estimating the Loss of Economic Predictability from Aggregating Firm-level Production Networks. *PNAS Nexus*, 3(3), 64.

Fan, H., Wu, T., & He, S. (2023). Research on the Industrial Chain Linkage Effect of Enterprise Digitalization. *China Industrial Economics(Zhongguo Gongye Jingji)*, (3), 115–132.

Fan, J. P. H., & Lang, L. H. P. (2000). The Measurement of Relatedness: An Application to Corporate Diversification. *Journal of Business*, 73(4), 629–660.

Lai, F., & Li, S. (2023). Common Ownership Networks and Innovation Spillovers of State-Owned Enterprises: An Empirical Study Based on Performance Evaluation Method Change. *Economic Research Journal(Jingji Yanjiu)*, (6), 119–136.

Lemelin, A. (1982). Relatedness in the Patterns of Inter-industry Diversification. *Review of Economics and Statistics*, 64(4), 646–657.

Li, L., & Wang, C. (2017). Tax Burden, Innovation Ability and Enterprise Upgrading: Empirical Evidence from NEEQ Listed Companies. *Economic Research Journal(Jingji Yanjiu)*, (11), 119–134.

Li, W., & Zheng, M. (2016). Is It Substantive Innovation or Strategic Innovation? Research on the Influence of Customer Concentration on the Enterprise Technological Innovation. *Economic Research Journal(Jingji Yanjiu)*, (4), 71–84.

Liu, W. (2022). Changes of Production Input Structure and Enterprise Innovation: An Analysis Based on the Endogenous Production Network. *Economic Research Journal(Jingji Yanjiu)*, (4), 50–67.

Lv, Y., Zhang, H., Xue, J., & Zhao, X. (2023). Do Tax Incentives Promote Emission Abatement of Firms: Evidence from China's Value-Added Tax Transformation Reform. *China Industrial Economics(Zhongguo Gongye Jingji)*, (2), 112–130.

Qian, X., Yang, Y., & Xu, W. (2010). Firm Network Position, Absorptive Capacity and Innovation Performance: An Interactive Effect Model. *Management World(Guanli Shijie)*, (5), 118–129.

Ren, B., & Hong, Y. (2005). An Approach of Improvement of the Economic Effectiveness in New-Type Industrialization: Analysis of the Visual Angle of a Kind of Industry Chain. *Journal of Northwest University (Philosophy and Social Sciences Edition)(Xibei Daxue Xuebao Zhexue Shehui Kexue Ban)*, (1), 47–54.

Wang, C., & Hu, Q. (2020). Knowledge Sharing in Supply Chain Networks: Effects of Collaborative Innovation Activities and Capability on Innovation Performance. *Technovation*, 94(3), 102010.

Wang, M., Chen, L., Zhang, Z., & Yao, Q. (2023). Collaborative Innovation Network Effect of Pilot Free Trade Zones: A Study Based on Spatial Regression Discontinuity and Geographical Identification. *The Journal of World Economy(Shijie Jingji)*, (3), 94–124.

Wang, R., Yan, B., & Deng, W. (2010). The Impact of FDI on the Independent Innovation Capability of Chinese Indigenous Industries: From the Perspective of Industrial Linkages. *China Industrial Economics(Zhongguo Gongye Jingji)*, (11), 16–25.

Yang, Y., Wei, J., & Luo, L. (2015). Who Innovate from the Government Subsidies? The Joint Moderate Effects of the Ownership and the Factor Market Distortions. *Management World(Guanli Shijie)*, (1), 75–86.

Zahra, S. A., & George, G. (2002). Absorptive Capacity: A Review, Reconceptualization, and Extension. *The Academy of Management Review*, 27(2), 185–203.

Zhang, C., Sun, P., & Xie, J. (2020). Does the Position of the Production Chain Affect the Choice of Innovation Modes? Theory and Evidence Based on Micro Perspective. *Management World(Guanli Shijie)*, (1), 45–59.

Zweimuller, J., & Brunner, J. K. (2005). Innovation and Growth with Rich and Poor Consumers. *Metroeconomica*, 56(2), 233–262.