

Original Research Article

Mechanisms and evidence of new quality agricultural productive forces empowering comprehensive grain production efficiency: A causal inference based on double machine learning

Xiaoding Tao

Anhui University of Finance and Economics, Bengbu, Anhui, 233030, China

Abstract: Do new quality agricultural productive forces actually move the needle on grain production efficiency in China? The old input-heavy growth model has run its course. We tackle this with panel data on 30 regions (2014–2022) and a DML estimator. The composite index lifts both land and labor productivity at 1% significance. Substantive pieces (labor, land, capital) and permeating ones (technology, data) complement, not crowd out. Three channels carry the load — Technological progress, resource reallocation, industrial upgrading. Effects hit hardest in major producing and balance areas, western provinces, and medium-development regions; high-level regions show diminishing returns. IV correction and swapping the learner leave the story unchanged.

Keywords: new quality agricultural productive forces; grain production efficiency; double machine learning

1. Introduction

Food security has been a pillar of China's strategy for decades. But around the mid-2010s, the old recipe — More fertilizer, more farmhands — Stopped paying off. Along came "new quality productive forces," which rearranges labor, capital, land, and know-how around innovation. Whether that rearrangement genuinely lifts grain production efficiency, and through what channels, is what we are after.

One wrinkle: These forces mix substantive inputs (upgraded land, modern equipment) with permeating capabilities like data-driven management. No single variable captures all of that. Another: Most prior work stays qualitative; hard causal evidence wrestling with confounders and endogeneity is thin on the ground.

We push three things forward. A composite index folding substantive and permeating sub-indicators into one score, traced to grain production efficiency. A look inside the black box at three transmission routes. And the DML framework from Zhang and Li (2023)^[7], which tames bias from high-dimensional confounders and reverse causality.

2. Theoretical analysis and research hypotheses

Four pillars underpin the concept: Human labor, cultivated land, capital stock, and scientific-technical knowledge. Each has been studied alone for years. What gets less attention is what happens when they upgrade together — Skilled operators on consolidated plots, modern machinery, data analytics guiding decisions. That bundling produces a qualitative jump, not just incremental gains. In measurement, the first three feed into the substantive tier (labor quality, land endowment, capital depth); scientific-technical knowledge spills into the permeating tier (technology penetration, informatization, data application) — Digital tools now seep into every traditional input.

That logic lands us at Hypothesis 1 (H1): New quality agricultural productive forces exert a measurable, positive effect on comprehensive grain production efficiency.

Hypothesis 2 (H2): Technological progress in agriculture serves as a significant transmission channel through which new quality productive forces lift grain production efficiency.

When information flows tighten and markets connect more deeply, labor, land, and capital find better homes. That eats into the misallocation drag plaguing Chinese agriculture.

Hypothesis 3 (H3): Resource allocation gains represent a second pathway through which new quality

agricultural productive forces promote grain production efficiency.

Rural economies now stretch into processing, logistics, e-commerce. Producers grab value downstream that used to slip past them.

Hypothesis 4 (H4): Industrial transformation and upgrading form a third conduit for the efficiency gains that new quality agricultural productive forces generate.

3. Data sources, research methods, and variable description

3.1. Data sources

We cover 30 provincial-level regions over 2014–2022. Tibet drops out for data gaps; Hong Kong, Macao, and Taiwan fall outside the mainland system. Numbers come from the China Statistical Yearbook, China Education Statistical Yearbook, China Rural Statistical Yearbook, and provincial yearbooks. Missing values are filled with linear interpolation.

3.2. Research methods

(1) Weighted Sum Method. The forces are multi-dimensional, so we roll sub-indicators into composite scores — One for substantive, one for permeating — Via a weighted linear model. Weights combine the entropy method and coefficient of variation, keeping things data-driven rather than subjective.

(2) Double Machine Learning. Endogeneity is real here: Productive forces and efficiency likely feed back into each other. Zhang and Li (2023) ^[7] proposed a DML setup using machine learning for nuisance estimation and orthogonalized moment conditions to isolate the causal parameter — Consistent and asymptotically normal under moderate assumptions.

The procedure nets out variation in treatment and outcome that controls explain, then feeds orthogonalized residuals into the final causal step. Cross-fitting holds overfitting in check; orthogonalization strips confounding. Standard errors absorb uncertainty from both stages.

Cross-fitting keeps overfitting at bay, and orthogonalization scrubs out confounding. As for standard errors, they absorb uncertainty from both estimation steps.

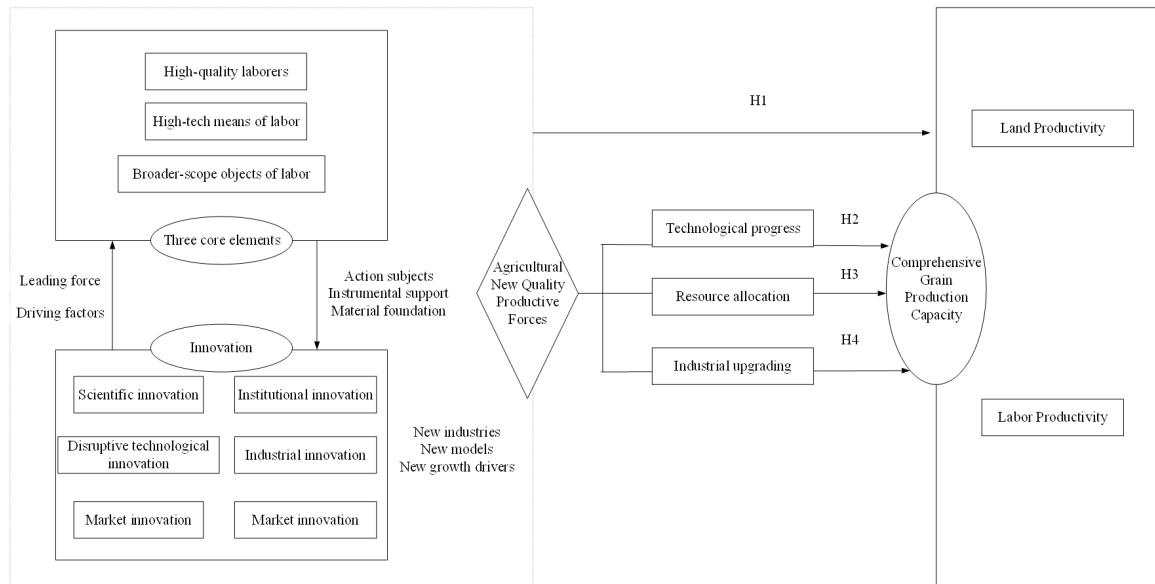


Figure 1. The theoretical logic of new quality agricultural productive forces driving the improvement of grain production efficiency.

3.3. Variable description

The dependent variable is comprehensive grain production efficiency, measured through land productivity and labor productivity following Yang Qing et al^[8]. Substantive elements (labor, land, capital) sit alongside permeating ones (science, technology, data), each built from several sub-indicators.

Two tiers make up the composite index. The substantive tier covers labor quality, land endowment, capital depth. The permeating tier spans technology penetration, informatization, data application. A weighted aggregation collapses these into one province-year score. Over 2014–2022 the national average

climbs, though provincial gaps are wide.

There is a risk that the aggregate masks offsetting forces. So we also pull the index apart into D1 (substantive) and D2 (permeating) and run each on its own.

(3) Eleven control variables come along to absorb provincial-level heterogeneity: Tax burden, economic development, social consumption, industrialization, urbanization, disaster rate, road mileage, freight volume, labor level, per capita fixed asset investment, and irrigation rate.

Summary statistics ($N = 300$): Y1 averages 4.845 (SD 2.618), Y2 3.888 (SD 1.693). The composite index D posts 0.170 (SD 0.079); D1 comes in at 0.195, D2 at 0.146. Among controls, economic development runs 11.044, urbanization 0.619, disaster rate 10.826, irrigation rate 0.442.

4. Empirical analysis

4.1. Baseline regression

Table 1 holds the baseline DML results. Columns (1)–(3) take land productivity; (4)–(6) take labor productivity. All carry linear and quadratic control terms plus year and province fixed effects. The headline: The productive forces index is positive and clears the 1% bar across the board.

Table 1. Baseline regression.

Variable Name	Land Productivity			Labor Productivity		
	(1)	(2)	(3)	(4)	(5)	(6)
Agricultural New Quality Productive Forces	16.607*** (4.503)			18.804*** (3.441)		
Substantive Elements		14.191*** (4.097)			16.068*** (3.602)	
Permeating Elements			18.749*** (4.943)			21.229*** (3.684)
First-order Terms of Control Variables	YES	YES	YES	YES	YES	YES
Second-order Terms of Control Variables	YES	YES	YES	YES	YES	YES
Year Fixed Effects	YES	YES	YES	YES	YES	YES
Province Fixed Effects	YES	YES	YES	YES	YES	YES
Observations	300	300	300	300	300	300

Decomposing in columns (2) and (5) for substantive elements and (3) and (6) for permeating ones, both come back positive and significant. The two sides are not trading off — Substantive inputs set the stage, permeating capabilities amplify the payoff. H1 survives.

4.2. Robustness test

Four robustness checks. IV estimation with mobile internet traffic — Still 1% significant. Winsorizing at 1% and 5%: Land 10.027 and 9.139, labor 7.591 and 6.208, all 1%. Dropping outlier provinces: 9.428 and 3.446. Swapping learner for Lasso (9.909, 8.309) and gradient boosting (7.468, 9.322). Coefficient stays positive throughout.

4.3. Mechanism analysis

What carries the effect? Technological progress posts 6.533 (1% significant). Resource reallocation shows up as mismatch reduction — Land mismatch drops 2.225, labor mismatch 1.510, both 1%. Industrial upgrading carries 0.945 (10% significant). Mediation tests back all three, lining up with H2, H3, and H4.

4.4. Heterogeneity analysis

Effects are not uniform. By producing-region: Significant in major producing (4.829, 10.368) and balance areas (9.384, 11.523), but land productivity in marketing areas misses (6.477). Geographically all three zones significant; land productivity smallest in central China (2.829), largest out west (15.107). By farm scale, labor productivity significant across all sizes, land only at moderate scale. By development level: Medium regions post the biggest coefficient (20.264); high-level regions already flatten (8.944, 10% only). Diminishing returns set in early where investment is advanced.

5. Conclusion and recommendations

The index tracks positively with land productivity (16.607) and labor productivity (18.804). Substantive (14.191, 16.068) and permeating (18.749, 21.229) components both pull weight. Three pathways — Technological progress (6.533), mismatch reduction (2.225 and 1.510), industrial upgrading (0.945) — Confirm H1–H4. Impact peaks in medium-level regions (20.264) and western China (15.107); high-level areas show diminishing returns. These survive endogeneity correction and all robustness checks.

Policy threads: Sustain R&D and extension spending; keep reforming land-transfer and credit markets; build infrastructure linking producers to downstream value. Since returns vary — Medium-level and western regions benefit most — Calibrating to local conditions beats one-size-fits-all.

Fundings

This research was funded by the Jiangxi Provincial Postdoctoral Research Project Merit-based Funding Program, grant number 2021KY48.

References

- [1] Jia Ru. Ensuring Food Security from the Perspective of the Big Food Concept: Interactive Mechanism, Logic Tracing, and Contemporary Significance [J]. *Journal of Northwest A&F University (Social Science Edition)*, 2024, 24(4): 63-73.
- [2] Gao Ming, Zhao Xue. Food Security from the Perspective of Building a Strong Agricultural Nation: Realistic Foundation, Challenges, and Promotion Strategies [J]. *Research of Agricultural Modernization*, 2023, 44(2): 185-195.
- [3] Wu Qi, Shi Kejing. Theoretical Analysis and Empirical Evidence of New Quality Productive Forces and Agricultural Economic Resilience [J]. *On Economic Problems*, 2025, (02): 111-120.
- [4] Luo Biliang. New Quality Productive Forces: Disruptive Innovation and Fundamental Transformation—Also on the Essential Provisions and Direction of Efforts for High-Quality Agricultural Development [J]. *China Rural Economy*, 2024(8).
- [5] Luo Biliang. On New Quality Agricultural Productive Forces [J]. *Reform*, 2024, (04): 19-30.
- [6] Yang Ying. Value Implications and Basic Ideas for Developing New Quality Agricultural Productive Forces [J]. *Issues in Agricultural Economy*, 2024, (04): 27-35.
- [7] Zhang Tao, Li Junchao. Network Infrastructure, Inclusive Green Growth, and Regional Disparities: A Causal Inference Based on Double Machine Learning [J]. *Journal of Quantitative & Technological Economics*, 2023, 40(4): 113-135.
- [8] Yang Qing, Jia Jiefei, Liu Jin, et al. How Do Agricultural Machinery Purchase Subsidies Affect Comprehensive Grain Production Capacity? [J]. *Management World*, 2023, 39(12): 106-121.