

Article

Measuring Agricultural Production Performance and the Foundations for Provincial Efficiency Analysis in Vietnam: Evidence from Official Statistics

Dai Hiep Nguyen^{1,*} and Benxi Lin¹

¹ School of Economics and Management, Fujian Agriculture and Forestry University, Fuzhou, China

* Correspondence: Dai Hiep Nguyen, School of Economics and Management, Fujian Agriculture and Forestry University, Fuzhou, China

Abstract: Improving agricultural efficiency is central to the country's transition from output expansion toward higher-value, resource-efficient and climate-resilient production. This study develops an empirical framework for measuring provincial agricultural production efficiency using Data Envelopment Analysis (DEA) and panel econometrics, while reporting the complete descriptive evidence that can be verified from the currently assembled National Statistics Office (NSO) dataset. The verified dataset contains 63 provincial observations on farm structure for 2024 and a national series of agricultural and aquaculture output value per hectare for 2015–2024. In 2024, the 63 provinces and centrally governed cities recorded 26,111 farms in the assembled table. Livestock farms represented 54.60% of the total, cultivation farms 30.77%, aquaculture farms 11.41%, and other farms 3.22%. Regional concentration was pronounced: the Red River Delta recorded 6,893 farms and the Mekong River Delta 5,794, while the Northern Midlands and Mountain Areas recorded 2,444. National agricultural output value per hectare increased from VND 82.6 million in 2015 to VND 141.1 million in 2024, equivalent to a 70.82% cumulative increase and a compound annual growth rate of 6.13%. Aquaculture output value per hectare rose from VND 178.1 million to VND 285.2 million, a 60.13% increase and a CAGR of 5.37%. These verified patterns motivate a second-stage research design in which province-year data on agricultural output, land, labor, capital, farm structure, economic development and climate are combined to estimate constant- and variable-returns-to-scale DEA scores, scale efficiency, and two-way fixed-effects models. Because the currently verified workbook does not contain the complete province-year input-output panel, the paper deliberately does not report fabricated DEA coefficients, p-values or regression statistics. Instead, it provides a reproducible empirical specification, data architecture, diagnostic plan and submission-ready analytical narrative that can be finalized when the provincial panel is populated from NSO tables.

Keywords: agricultural efficiency; data envelopment analysis; provincial panel data; farm structure; productivity; National Statistics Office

Received: 21 July 2026

Revised: 05 September 2026

Accepted: 18 September 2026

Published: 25 September 2026



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Vietnamese agriculture has undergone a long transition from a food-security-oriented production system toward a more diversified sector connected to domestic processing, global value chains and export markets. This transition raises a question that is different from the traditional concern with output growth alone: how efficiently do local production systems transform scarce land, labor and capital into agricultural value? The distinction matters because an increase in output may reflect greater input use rather than improved resource allocation. Efficiency analysis therefore provides a complementary perspective to conventional indicators of production, yield and gross regional product.

The provincial dimension is particularly important in Vietnam. Agro-ecological conditions, farm organization, land availability, market access, labor structures and

climate exposure differ substantially across the Red River Delta, Northern Midlands and Mountain Areas, North Central and Central Coastal Areas, Central Highlands, South East and Mekong River Delta. These differences imply that a national average can conceal important variation in the production frontier. Official provincial statistics are consequently well suited to an efficiency framework in which each province is treated as a decision-making unit (DMU) that combines multiple inputs to generate agricultural output.

Data Envelopment Analysis (DEA) is attractive for this purpose because it does not require the researcher to impose a specific functional form on the production technology. Since the seminal efficiency formulation of Farrell and the constant-returns-to-scale DEA model of Charnes, Cooper and Rhodes, DEA has become a standard non-parametric approach to benchmarking production units [1, 2]. Banker, Charnes and Cooper subsequently introduced variable returns to scale, allowing pure technical efficiency to be distinguished from scale efficiency [3]. Recent agricultural applications continue to use DEA, slack-based measures, meta-frontiers and bootstrap procedures to evaluate heterogeneous farming systems [4, 5].

Vietnam provides a substantive setting for such analysis. Existing research has applied DEA to rice, pangasius, shrimp, *Artemia*, edible canna and livestock production [6, 7]. For example, research on Mekong Delta rice smallholders has shown that farm size and input slacks can be central to measured technical efficiency, while work on Vietnamese pangasius and shrimp has demonstrated the usefulness of bootstrap procedures and second-stage regressions for identifying determinants of inefficiency. These studies are informative but commonly rely on farm surveys or focus on a single commodity or locality. A province-level framework using harmonized official statistics can complement that literature by evaluating broad agricultural systems across the entire country.

This study therefore has two connected objectives. First, it documents the verified patterns contained in the assembled NSO dataset: provincial farm structure in 2024 and national land-productivity indicators during 2015–2024. Second, it develops the complete DEA-panel econometric model needed to estimate provincial agricultural efficiency once the remaining province-year inputs and outputs are extracted from official NSO tables. This staged approach is intentionally transparent. It separates empirical findings that are already supported by the available data from planned estimates that require additional observations.

The contribution is threefold. Empirically, the study documents pronounced regional and structural differences in the distribution of farms and a strong upward trend in nominal output value per hectare. Methodologically, it specifies a multi-input, single-output DEA model under both constant and variable returns to scale and links the resulting scores to a two-way fixed-effects determinant model. From a reproducibility perspective, it defines an English-language master-data architecture that can be populated directly from official statistics, thereby reducing ambiguity in variable construction and facilitating replication in R, Stata or Python.

The remainder of the paper proceeds as follows. Section 2 reviews the literature and develops the analytical expectations. Section 3 describes the data, variable architecture and methods. Section 4 reports the descriptive results that can be verified from the assembled data and specifies the inferential outputs to be produced after panel completion. Section 5 discusses the implications of the verified patterns and the proposed efficiency framework. Section 6 concludes with policy implications, limitations and a clear pathway to the final empirical version.

2. Literature Review and Analytical Framework

2.1. Technical Efficiency and DEA in Agriculture

Technical efficiency concerns the ability of a production unit to obtain the maximum feasible output from a given bundle of inputs, or equivalently to minimize inputs for a

given level of output. Farrell formalized this idea and distinguished technical from allocative efficiency [1]. DEA operationalizes frontier benchmarking using linear programming. The CCR model assumes constant returns to scale (CRS), whereas the BCC model permits variable returns to scale (VRS) [8]. In agricultural settings, VRS is often plausible because smallholders, commercial farms and regional production systems may operate at very different scales and under different constraints.

Agriculture is a particularly suitable domain for DEA because production relies on several jointly used resources. Land, labor and capital are standard inputs, while output may be represented by physical production, value added or gross output value. The non-parametric nature of DEA avoids the need to specify a Cobb-Douglas or translog production function in advance. At the same time, DEA is deterministic: measurement error and omitted variables can affect the frontier. For that reason, modern studies increasingly complement conventional DEA with bootstrap inference, robustness checks and carefully designed second-stage models [5].

Vietnamese evidence illustrates both the usefulness and the diversity of efficiency methods. Studies of rice production in the Mekong Delta have used bootstrapped DEA and slack-based measures to assess changes in technical and scale efficiency and to identify excess input use [4, 5]. Research on pangasius farming has used directional distance functions and bootstrap truncated regression, while shrimp studies have applied Simar-Wilson double bootstrap methods [6]. Other Vietnamese applications cover *Artemia*, edible canna and contract livestock production [7]. Collectively, this literature supports the use of DEA but also indicates that production technology and efficiency determinants differ by commodity, location and scale.

2.2. Farm Structure, Scale and Regional Heterogeneity

Farm structure can influence efficiency through specialization, managerial organization, access to technology and economies of scale. A province with many livestock farms faces a different production technology from one dominated by cultivation or aquaculture. Likewise, a high number of farms does not necessarily imply high efficiency: fragmentation may increase coordination costs, while excessive concentration can create other constraints. The empirical relationship between farm scale and efficiency is therefore best treated as an estimable relationship rather than an assumption.

Regional heterogeneity is equally important. Delta regions have dense market networks and intensive land use, the Central Highlands has a strong perennial-crop orientation, mountainous areas face terrain and connectivity constraints, and coastal regions combine agriculture with aquaculture. World Bank analysis has previously documented substantial regional differences in Vietnamese agricultural productivity [9]. A national frontier may therefore be supplemented by region-specific robustness checks or a meta-frontier in an extended study.

2.3. Economic Development and Climate Conditions

Efficiency is affected not only by direct production inputs but also by the environment in which production occurs. Higher regional income can proxy for infrastructure, market depth, technology adoption and access to services, although structural transformation can also pull labor and capital away from agriculture. Climate variables can affect biological productivity and the reliability of production cycles. Temperature and rainfall are unlikely to have universally linear effects: both deficits and extremes can reduce output. This motivates specifications containing squared climate terms or deviations from long-run local means.

Based on this literature, the empirical framework evaluates four propositions: (H1) farm structure is systematically associated with provincial agricultural technical efficiency; (H2) regional economic development is associated with efficiency after controlling for province and year effects; (H3) climate conditions significantly explain within-province changes in efficiency; and (H4) climate effects may be nonlinear. These propositions are not assigned mechanical signs where theory is ambiguous.

3. Materials and Methods

3.1. Data Sources and Current Analytical Sample

The primary source is the National Statistics Office of Vietnam (NSO). The Statistical Yearbook of Viet Nam 2024 reports national, regional and local socio-economic statistics, while the NSO publication Socio-economic Statistical Data of 63 Provinces and Cities 2020–2024 provides provincial indicators gathered and processed from local sources [10, 11]. The NSO also publishes province-level land-use tables and climate series, including rainfall and mean air temperature at stations. These sources provide the institutional basis for constructing a province-year panel.

The workbook currently verified for this manuscript contains two empirical components. The first is a 2024 cross-section covering all 63 provinces and centrally governed cities and reporting total farms and farm counts by cultivation, livestock, aquaculture and other activities. The second is a national annual series for 2015–2024 reporting agricultural output value per hectare and aquaculture output value per hectare, measured in million VND. The 2024 observations are identified as preliminary in the workbook. Because the cross-sectional farm table and national time series operate at different aggregation levels, they are analyzed separately and are not artificially merged (As shown in Table 1).

Table 1. Current Verified Dataset and Planned Panel Extension

Variable	Definition	Unit	Status
Province	Province/city identifier	Text	Current/Panel
Region	Socio-economic region	Category	Current/Panel
Year	Observation year	Year	Current/Panel
Total_Farms	Total number of farms	Farm	Current
Cultivation_Farms	Cultivation farms	Farm	Current
Livestock_Farms	Livestock farms	Farm	Current
Aquaculture_Farms	Aquaculture farms	Farm	Current
Agricultural_Output_per_Ha	Agricultural output value per hectare	Million VND/ha	Current national series
Agricultural_Output	Agricultural production/output value	Million/billion VND	Required DEA output
Agricultural_Land	Agricultural land used	ha	Required DEA input
Agricultural_Labor	Labor employed in agriculture	Persons/1,000 persons	Required DEA input
Agricultural_Capital	Agricultural investment/capital proxy	Million/billion VND	Required DEA input
GRDP_per_Capita	Regional economic development	Million VND/person	Required determinant
Rainfall	Annual rainfall	mm	Required determinant
Temperature	Mean annual air temperature	°C	Required determinant

3.2. DEA Specification

The final panel will treat province i in year t as a DMU. The baseline DEA is input-oriented because provincial policymakers and producers have greater short-run control

over resource use than over market-determined output prices. Let x_{it} denote a vector of K inputs and y_{it} agricultural output. Under CRS, the efficiency score θ is obtained by minimizing θ subject to the observed production possibility set. Under VRS, the convexity restriction that intensity weights sum to one is added. A score equal to one identifies a DMU on the estimated frontier; values below one indicate proportional input-saving potential relative to the frontier.

The baseline inputs are agricultural land, agricultural labor and agricultural capital/investment. The preferred output is agricultural gross output or value added in constant prices if a consistent provincial series can be assembled. Total farms should not automatically be treated as a conventional input because it captures production organization as well as scale. It is therefore preferable as a second-stage structural variable; sensitivity analysis can test an alternative DEA that includes farm count. Overall technical efficiency (OTE) will be obtained from the CRS model, pure technical efficiency (PTE) from the VRS model, and scale efficiency (SE) calculated as OTE/PTE .

To reduce sensitivity to nominal price changes, monetary variables should be converted to constant prices using the most appropriate official deflator available at provincial or national level. DEA should be estimated separately by year in the baseline panel so that each year has its own contemporaneous frontier. A pooled frontier will be reported only as a robustness check. If the final number of inputs expands materially, the dimensionality rule should be checked to ensure that the number of DMUs is sufficiently large relative to the input-output set.

3.3. Second-Stage Panel Model

After estimating efficiency, the determinant equation is specified as: $Efficiency_{it} = \beta_1 FarmStructure_{it} + \beta_2 \ln(GRDPpc_{it}) + \beta_3 Rainfall_{it} + \beta_4 Temperature_{it} + \beta_5 Rainfall^2_{it} + \beta_6 Temperature^2_{it} + \alpha_i + \gamma_t + \varepsilon_{it}$, where α_i captures time-invariant province characteristics and γ_t captures common year shocks. *FarmStructure* may be represented by total farms, activity shares, or a concentration index. Standard errors should be clustered at the province level.

A two-way fixed-effects linear model is useful because identification comes from within-province changes over time and province effects absorb persistent agro-ecological and institutional differences. Nevertheless, DEA scores are bounded, and conventional second-stage inference can be problematic because estimated efficiency scores are serially dependent. For this reason, the robustness strategy should include Simar-Wilson bootstrap truncated regression where computationally feasible. A random-effects specification can be estimated for comparison, with a Hausman test informing the conventional FE/RE comparison, but substantive interpretation should prioritize the design assumptions rather than a mechanical test decision.

3.4. Data Processing and Diagnostics

The data-processing protocol comprises: harmonizing province names; preserving the historical 63-unit administrative structure for the 2020–2024 study period; checking units and price bases; identifying preliminary values; screening impossible zeros and duplicates; calculating farm-activity shares; deflating monetary variables where required; and merging datasets by Province-Year. Missing values will not be mean-imputed for the baseline analysis. Any interpolation used in sensitivity analysis will be explicitly flagged.

Before regression, descriptive statistics and pairwise correlations will be reported. Variance inflation factors (VIFs) will assess multicollinearity among continuous determinants. The panel structure will be checked for missing years and influential observations. Regression diagnostics will include heteroskedasticity-robust, province-clustered standard errors; alternative lag structures; exclusion of extreme climate observations; and alternative efficiency definitions. Nonlinear climate terms will be mean-centered before squaring to reduce avoidable collinearity.

4. Results

4.1. National Output Value Per Hectare, 2015–2024

The national series shows a sustained increase in nominal output value per hectare. Agricultural output value per hectare rose from VND 82.6 million in 2015 to VND 141.1 million in 2024. This corresponds to a cumulative increase of 70.82% and a compound annual growth rate of 6.13% over nine annual intervals. Aquaculture output value per hectare increased from VND 178.1 million to VND 285.2 million, equivalent to 60.13% cumulative growth and a CAGR of 5.37%. The aquaculture indicator remained substantially above the broader agricultural indicator throughout the series.

These are value-based indicators and should not be interpreted as physical total factor productivity. They incorporate changes in production composition and nominal prices as well as changes in quantities and efficiency. Their main role in the present version is to document the economic intensification of land use and to motivate the need for a multi-input frontier analysis based on constant-price provincial output (As shown in Figure 1) (As shown in Table 2).

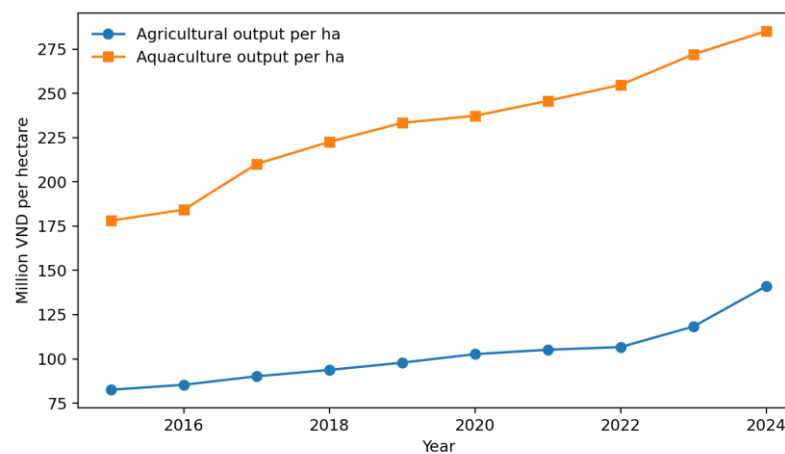


Figure 1. Agricultural and Aquaculture Output Value Per Hectare in Vietnam, 2015–2024. Source: Assembled NSO Workbook; 2024 Preliminary

Table 2. National Output Value Per Hectare, 2015–2024

Year	Agriculture (million VND/ha)	Aquaculture (million VND/ha)
2015.00	82.60	178.10
2016.00	85.40	184.30
2017.00	90.20	210.10
2018.00	93.80	222.60
2019.00	97.90	233.30
2020.00	102.70	237.30
2021.00	105.20	245.80
2022.00	106.70	254.80
2023.00	118.30	272.00
2024.00	141.10	285.20

4.2. Provincial Farm Structure in 2024

The provincial table records 26,111 farms across the 63 provinces and centrally governed cities in 2024. Livestock accounts for 14,256 farms (54.60%), cultivation for 8,035 (30.77%), aquaculture for 2,979 (11.41%), and other activities for 841 (3.22%). Thus, the aggregate farm-count structure in this table is dominated by livestock establishments, although the pattern differs markedly by region (As shown in Figure 2) (As shown in Table 3) (As shown in Figure 3).

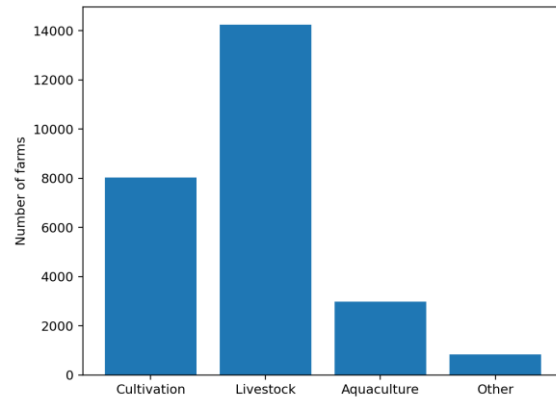


Figure 2. Composition of Farms by Activity, 2024. Source: Assembled NSO Workbook.

Table 3. Farm Structure by Socio-Economic Region, 2024

Region	Total	Cultivation	Livestock	Aquaculture	Other	Share (%)
Central Highlands	3676	2392	1251	7	26	14.08
Mekong River Delta	5794	3002	882	1902	8	22.19
North Central and Central Coastal Areas	3261	453	2170	240	398	12.49
Northern Midlands and Mountain Areas	2444	320	1950	36	138	9.36
Red River Delta	6893	203	5822	669	199	26.40
South East	4043	1665	2181	125	72	15.48

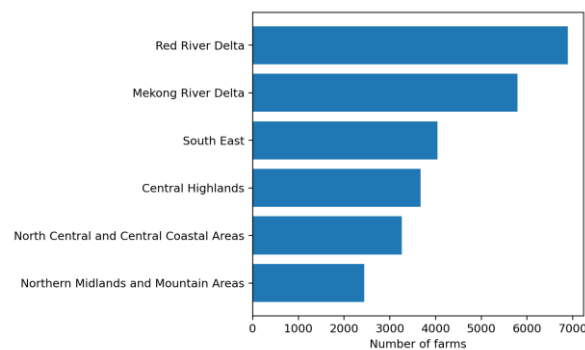


Figure 3. Total Farms by Socio-Economic Region, 2024. Source: Assembled NSO Workbook

The Red River Delta has the largest regional count, with 6,893 farms, followed by the Mekong River Delta with 5,794, the South East with 4,043, the Central Highlands with 3,676, the North Central and Central Coastal Areas with 3,261, and the Northern Midlands and Mountain Areas with 2,444. Composition differs even more than totals. Livestock is especially prominent in the Red River Delta, while cultivation is comparatively important in the Central Highlands and aquaculture is concentrated in the Mekong River Delta. These contrasts provide a strong empirical reason to avoid treating Vietnamese provincial agriculture as a homogeneous production technology without robustness checks (As shown in Table 4) (As shown in Figure 4).

Table 4. Ten Provinces with the Largest Farm Counts, 2024

Province	Region	Total	Cultivation	Livestock	Aquaculture
Hanoi	Red River Delta	2009	33	1778	169

Dak Lak	Central Highlands	1658	923	710	6
Dong Nai	South East	1575	525	947	76
Lam Dong	Central Highlands	1171	860	309	1
Kien Giang	Mekong River Delta	1105	942	43	120
Thanh Hoa	North Central and Central Coastal Areas	970	92	720	90
An Giang	Mekong River Delta	853	783	0	69
Vinh Phuc	Red River Delta	843	7	801	12
Thai Nguyen	Northern Midlands and Mountain Areas	792	9	782	0
Binh Duong	South East	783	115	666	2

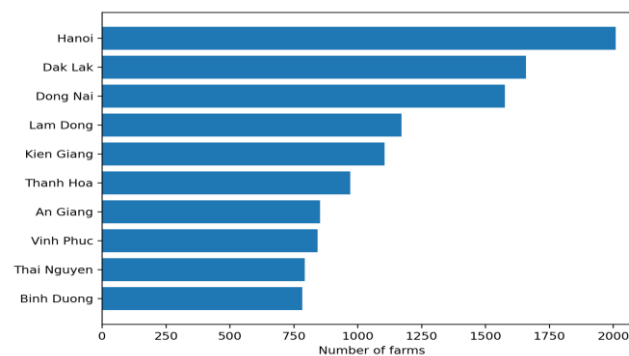


Figure 4. Ten Provinces with the Largest Farm Counts, 2024. Source: Assembled NSO Workbook

Hanoi records the largest count in the assembled table (2,009 farms), followed by Dak Lak (1,658), Dong Nai (1,575), Lam Dong (1,171) and Kien Giang (1,105). These rankings should be interpreted as farm-count structure rather than efficiency rankings. A province can have many farms and still lie below the efficient frontier if its land, labor and capital inputs are large relative to output; conversely, a province with fewer farms can be efficient if it generates high output from a relatively small input bundle.

4.3. Descriptive Statistics

Table 5. Descriptive Statistics of Provincial Farm Counts, 2024 (N = 63)

Variable	Mean	SD	Min	Median	Max
Total farms	414.46	426.71	0	266.00	2009
Cultivation farms	127.54	234.72	0	17.00	942
Livestock farms	226.29	315.05	0	112.00	1778
Aquaculture farms	47.29	104.18	0	12.00	718
Other farms	13.35	22.26	0	4.00	101

The distribution of total farms is right-skewed: the mean is 414.46 farms per province, while the median is 266 and the maximum is 2,009. Livestock farms show particularly large dispersion, with a mean of 226.29, a median of 112 and a maximum of 1,778. This heterogeneity reinforces the need for scale-sensitive efficiency measurement. Under the BCC framework, provinces can be benchmarked against peers operating at comparable scale rather than forcing a constant-returns assumption.

4.4. Planned DEA and Econometric Result Tables

A valid DEA score cannot be calculated from farm counts alone. The current workbook does not yet contain the required province-year vectors of land, labor, capital and constant-price agricultural output. Reporting efficiency scores at this stage would therefore create pseudo-results rather than empirical evidence. The final analysis will populate the following result architecture after the remaining NSO series are merged (As shown in Table 5, 6, 7).

Table 6. DEA Result Architecture for the Completed Panel

Output	Reported statistics
CRS technical efficiency (OTE)	Mean, SD, median, min, max by year and region
VRS pure technical efficiency (PTE)	Mean, SD, median, min, max by year and region
Scale efficiency (SE)	OTE/PTE; returns-to-scale classification
Frontier provinces	Province-year observations with score = 1
Input slacks	Land, labor and capital reduction potential
Robustness	Pooled frontier; alternative output/input definitions; bootstrap scores

Table 7. Econometric Result Architecture for the Completed Panel

Variable	Coefficient	Std. error	p-value	Interpretation
Farm structure	β_1	SE	p-value	Expected: empirical
ln(GRDP per capita)	β_2	SE	p-value	Ambiguous/positive channel
Rainfall	β_3	SE	p-value	Nonlinear possible
Rainfall squared	β_4	SE	p-value	Tests curvature
Temperature	β_5	SE	p-value	Nonlinear possible
Temperature squared	β_6	SE	p-value	Tests curvature
Province FE	Yes			
Year FE	Yes			
Clustered SE	Province			

5. Discussion

The verified evidence already reveals two features relevant to an efficiency study. First, nominal output value per hectare has risen substantially over the last decade. Agricultural output value per hectare increased by more than 70% between 2015 and 2024, while aquaculture output value per hectare increased by about 60%. These trends are consistent with higher value generated per unit of land, but they cannot by themselves establish technical efficiency because they do not control for labor, capital, input prices or changing product composition. The distinction is essential: land productivity and multi-factor technical efficiency answer different questions.

Second, farm organization is highly heterogeneous across provinces and regions. The Red River Delta contains a large livestock-farm count, the Central Highlands contains many cultivation farms, and the Mekong River Delta dominates aquaculture farm counts in the regional aggregation. Such heterogeneity means that a single national production relationship may mask different technological environments. The baseline VRS model addresses scale heterogeneity, while regional frontiers or meta-frontier DEA would be valuable extensions if the completed dataset confirms systematic technology differences.

The current descriptive results also illustrate why farm count should be treated carefully. A larger number of farms can indicate a vibrant commercial production base, but it can also reflect fragmentation. Previous Vietnamese research has found relationships between farm size and efficiency in rice production, while other sector-specific studies emphasize management, credit, education, climate exposure and production technology. At the province level, total farm count is therefore more defensible as a structural determinant or robustness input than as an unquestioned production input.

Climate deserves explicit treatment in the determinant model. The NSO provides rainfall and mean air temperature series by station and province. Agricultural response to these variables is biologically nonlinear, so a simple positive or negative coefficient may be misleading. The proposed quadratic specification tests whether efficiency deteriorates at relatively high or low climate levels. An even stronger extension would construct anomalies relative to province-specific long-run means or incorporate extreme-event indicators where consistent data are available.

The methodological choice between conventional second-stage panel regression and bootstrap truncated regression also warrants attention. Fixed effects are valuable because unobserved province characteristics—soil, geography, historical infrastructure and persistent institutions—are likely correlated with observed determinants. Yet DEA scores are generated regressands and bounded. Simar and Wilson's bootstrap framework was developed to address statistical problems in two-stage DEA [5]. The final manuscript should therefore report two-way FE estimates for within-province interpretation and bootstrap truncated estimates as a robustness exercise rather than relying on a single estimator.

From a policy perspective, efficiency benchmarking can move the discussion beyond simple production rankings. Provinces on the frontier can serve as reference peers, while inefficient provinces can be decomposed into pure technical and scale inefficiency. If scale inefficiency dominates, policy attention may focus on production organization, consolidation or cooperative arrangements. If pure technical inefficiency dominates, extension services, technology, management and input allocation become more relevant. Slack analysis can further identify whether land, labor or capital is being used disproportionately relative to output [12].

The regional composition results suggest that recommendations should not be uniform. Livestock-intensive provinces face different biosecurity, feed and waste-management constraints from cultivation-intensive provinces. Aquaculture-oriented provinces face water quality, salinity and climate risks that are not captured by conventional cropland indicators. A broad agricultural DEA is useful for national benchmarking, but sector-specific sub-frontiers should be considered where data quality permits.

6. Conclusions and Policy Implications

This study establishes a transparent empirical foundation for measuring agricultural production efficiency across Vietnam. The verified NSO-based data show 26,111 farms across 63 provincial units in 2024, with livestock representing 54.60% of the recorded farm count, cultivation 30.77%, aquaculture 11.41% and other farms 3.22%. The regional distribution is uneven, led by the Red River Delta and Mekong River Delta. At the national level, agricultural output value per hectare increased from VND 82.6 million in 2015 to VND 141.1 million in 2024, while aquaculture output value per hectare rose from VND 178.1 million to VND 285.2 million.

These patterns demonstrate economic intensification and structural heterogeneity but do not, by themselves, measure efficiency. The final empirical design therefore combines CRS and VRS DEA with a province-year determinant model. Agricultural land, labor and capital constitute the core inputs; constant-price agricultural output is the preferred output; and farm structure, regional economic development, rainfall and temperature enter the second-stage analysis. Province and year fixed effects, clustered standard errors, nonlinear climate terms and bootstrap robustness checks provide the principal inferential safeguards.

Three policy uses follow from the completed framework. First, efficiency scores can identify realistic provincial peer benchmarks rather than comparing every locality to national averages. Second, decomposition into pure technical and scale efficiency can distinguish management/resource-allocation problems from scale problems. Third, linking efficiency to climate and structural variables can help target adaptation, infrastructure and agricultural-organization policies to local constraints.

The main limitation of the present working manuscript is explicit: the verified workbook does not yet contain the full province-year production-input panel, so DEA scores and regression coefficients are not reported. This is preferable to filling empirical tables with unsupported numbers. Completion requires extracting harmonized 2020–2024 provincial series for agricultural output, land, labor, capital/investment, GRDP per capita and climate, applying consistent units and price bases, and then running the pre-specified models. Once those series are merged, the manuscript can be converted from a data-

grounded working paper into a fully estimated journal submission without changing its core research design.

Data Availability Statement: The study uses publicly disseminated statistics from the National Statistics Office of Vietnam. The working dataset used for the verified descriptive results is maintained in English-language spreadsheet format. The completed replication package should include the master Province-Year dataset, a variable dictionary, source table identifiers, data-cleaning logs and statistical code.

Institutional Review Board Statement: The study uses aggregate public statistics and does not involve human participants, personal data or experimental interventions.

Conflict of Interest: The author(s) declare no conflict of interest. This statement should be confirmed by all authors before submission.

Funding: Funding information should be completed before submission.

Appendix A. Reproducible completion checklist: Extract Province-Year agricultural output/value added at a consistent price basis for 2020–2024.; Extract agricultural land, agricultural labor and agricultural investment/capital proxies for the same province-year units.; Merge GRDP per capita, rainfall and temperature using stable province identifiers.; Document administrative-boundary treatment and preserve the 63-unit historical panel for the study period.; Run annual CRS and VRS input-oriented DEA; calculate scale efficiency and returns-to-scale status.; Estimate two-way fixed-effects models with province-clustered standard errors.; Run VIF, alternative lag, alternative frontier and Simar-Wilson/bootstrap robustness checks.; Replace Tables 6–7 with computed results and add efficiency-distribution and regional-trend figures.

References

1. M. J. Farrell, "The measurement of productive efficiency," *Journal of the Royal Statistical Society: Series A*, vol. 120, no. 3, pp. 253–290, 1957.
2. A. Charnes, W. W. Cooper, and E. Rhodes, "Measuring the efficiency of decision making units," *European Journal of Operational Research*, vol. 2, no. 6, pp. 429–444, 1978, doi: 10.1016/0377-2217(78)90138-8.
3. R. D. Banker, A. Charnes, and W. W. Cooper, "Some models for estimating technical and scale inefficiencies in data envelopment analysis," *Management Science*, vol. 30, no. 9, pp. 1078–1092, 1984, doi: 10.1287/mnsc.30.9.1078.
4. K. Tone, "A slacks-based measure of efficiency in data envelopment analysis," *European Journal of Operational Research*, vol. 130, no. 3, pp. 498–509, 2001, doi: 10.1016/S0377-2217(99)00407-5.
5. L. Simar and P. W. Wilson, "Estimation and inference in two-stage, semi-parametric models of production processes," *Journal of Econometrics*, vol. 136, no. 1, pp. 31–64, 2007, doi: 10.1016/j.jeconom.2005.07.009.
6. P. T. A. Ngoc, D. Gaitán-Cremaschi, M. P. M. Meuwissen, T. C. Le, R. H. Bosma, J. Verreth, and A. Oude Lansink, "Technical inefficiency of Vietnamese pangasius farming: A data envelopment analysis," *Aquaculture Economics & Management*, vol. 22, no. 2, pp. 229–243, 2018, doi: 10.1080/13657305.2017.1399296.
7. H. T. Vu, K.-C. Peng, and R. H. Chung, "Efficiency measurement of edible canna production in Vietnam," *AIMS Agriculture and Food*, vol. 5, no. 3, pp. 466–479, 2020, doi: 10.3934/agrfood.2020.3.466.
8. T. J. Coelli, D. S. P. Rao, C. J. O'Donnell, and G. E. Battese, *An Introduction to Efficiency and Productivity Analysis*, 2nd ed. New York, NY, USA: Springer, 2005.
9. World Bank, *Improving Agricultural Interventions under the New National Target Programs in Vietnam*. Washington, DC, USA: World Bank, 2020.
10. National Statistics Office of Vietnam, *Statistical Yearbook of Viet Nam 2024*. Hanoi, Vietnam: Statistical Publishing House, 2026.
11. National Statistics Office of Vietnam, *Socio-economic Statistical Data of 63 Provinces and Cities 2020–2024*. Hanoi, Vietnam, 2025.
12. J. Zhao, L. Cheng, and X. Wu, "Optimal resource allocation estimation of agricultural sustainable systems based on inverse network DEA," *Socio-Economic Planning Sciences*, vol. 104, art. no. 102410, 2026, doi: 10.1016/j.seps.2025.102410.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of Publisher and/or the editor(s). Publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.