The Efficiency Layer of a Performance Evaluation Model: Data Envelopment Analysis of Artificial-intelligence Investment in Chinese E-commerce Platforms

Guojin Lin*
Farabi International Business School, Al-Farabi Kazakh National University, Almaty 050040, Republic of Kazakhstan
*Corresponding email: lgj3089@gmail.com

Abstract

Composite performance scores measure the level of a firm's transformation but not the efficiency with which resources are converted into it. This paper constructs the efficiency layer of the Pathway-Capability-Performance (PCP) framework by applying data envelopment analysis to a panel of six platform-years assembled from the results announcements of Alibaba Group, JD.com, and PDD Holdings. Research and development (R&D) expenditure is the single input, revenue and net income attributable to ordinary shareholders are the outputs. Under an input-oriented, constant-returns specification, two units lie on the frontier and four fall below it, with scores ranging from 0.299 to 1.000. Efficiency declined at all three platforms between their two most recent reported fiscal years, and this decline persists across every alternative specification tested: a non-Generally Accepted Accounting Principles (GAAP) profit measure, a revenue-only single-output model, and variable returns to scale. Its mechanism is transparent - research spending grew faster than revenue at all three platforms, while profit fell at all three. The ranking of the declines, by contrast, is not robust. It reverses across specifications, and the paper therefore reports it as specification-dependent rather than as a finding, correcting an interpretation advanced in an earlier version of this analysis. A scale decomposition locates almost all of Alibaba's measured inefficiency in scale rather than in pure technical efficiency, which is consistent with - though it does not identify - a build-out interpretation. The paper documents three constraints under which frontier analysis operates in concentrated markets: units below conventional thresholds, censoring of frontier units in longitudinal use, and the inseparability of a market-wide shock from firm-specific effects. It also specifies an eight-point protocol for reporting the efficiency layer so that these constraints remain visible to users of the model.

Keywords

Data envelopment analysis, Concentrated market, E-commerce, Platform

Introduction

A composite performance score indicates how far a firm has progressed, but not how efficiently it has progressed. Two platforms may reach similar positions in a technological transition while consuming very different resources; for anyone allocating capital, this difference matters more than the position itself. This paper constructs the component of the Pathway-Capability-Performance (PCP) framework that addresses this question: its efficiency layer.

The framework models large-language-model reconstruction as a sequence in which implementation pathways reshape capabilities, and the reshaped capabilities generate performance across five dimensions. This follows the broader treatment of digital transformation as a process linking strategic choices to organizational resources and outcomes [1]. Composite scoring, developed separately, sits at the end of that sequence. Efficiency benchmarking sits alongside it, asking what each unit of resource purchased achieves. Data envelopment analysis (DEA) is the natural instrument for this, as it requires no assumption about functional form and accommodates multiple outputs [2,3]. Composite PCP scoring and efficiency benchmarking are complementary rather than alternative:

The first records a position, the second records the cost of reaching it, and the framework requires both.

Applying the instrument in a concentrated platform market raises difficulties that this paper treats as its subject rather than as caveats appended to a result. Three platforms dominate Chinese e-commerce, and even pooling data across reporting years yields few observations relative to the dimensions of any plausible specification. Three consequences follow. The estimator's discriminating power is weak, so units may be scored as efficient by construction rather than by performance. Units already on the frontier cannot be observed to improve, so longitudinal comparison beginning from the frontier is censored from above. Furthermore, because every unit is drawn from one market in one period, a market-wide change in trading conditions cannot be separated from firm-specific effects. The paper's contribution is accordingly threefold. First, it reports an application in which the full set of defensible specifications is estimated and disclosed, rather than a single preferred specification with the others suppressed. Second, it shows that at this sample size the direction of efficiency change is identified robustly while the ranking of changes across firms is not - a distinction that matters precisely because the ranking is the quantity a reader is most tempted to interpret. Third, it specifies a reporting protocol for the efficiency layer that keeps these limits visible to users of the framework.

Method and the constraints of concentrated markets

Frontier estimation

DEA constructs a piecewise-linear frontier from observed units and measures each unit's radial distance from it. For unit o the input-oriented constant-returns-to-scale (CRS) programme minimises θ subject to $\Sigma\lambda_j x_j \leq \theta x_o$, $\Sigma\lambda_j y_j \geq y_o$ and $\lambda \geq 0$. Adding the convexity constraint $\Sigma\lambda_j = 1$ yields the variable-returns (VRS) model of Banker, Charnes. The CRS score measures technical efficiency inclusive of scale effects; the VRS score measures pure technical efficiency; and their ratio, $SE=\theta\_CRS/\theta\_VRS$, is the unit's scale efficiency. When a unit achieves a high VRS score yet records a low CRS score, it performs well given its existing scale of operation but operates at a scale that cannot be efficiently transformed by the observed technology; this distinction forms the primary basis for subsequent interpretation.

Discrimination and the number of units

The method's discriminating power depends on the ratio of units to dimensions. When units are few relative to the number of inputs and outputs, many are projected onto the frontier by construction and scored as efficient, whether or not they truly are. Two conventions address this issue. Golany and Roll propose that the number of units should be at least twice the sum of inputs and outputs [4]. Dyson et al., documenting the problem among a wider set of pitfalls, proposed at least three times that sum [5]. Cook et al. added two requirements that bear on design rather than sample size: that model choice should precede inspection of the data, and that homogeneity among units is a precondition rather than a convenience [6].

All three cautions bear directly on the present application. The specification adopted for this analysis was fixed prior to estimation. It exactly satisfies the Golany-Roll threshold but falls three units short of the Dyson threshold. Furthermore, it rests on a homogeneity assumption that is imperfect on both the input and output sides. Each of these points is stated explicitly in the specification rather than discovered during the discussion.

Censoring, common shocks, and inference

Three additional constraints apply to short panels comprising few units and are less frequently noted. The first is censoring. An efficiency score is bounded above by unity, so a unit that defines the frontier in the first period cannot be observed to improve in the second. If the score of such a unit changes, the change can only be a decline; if it does not change, no inference about improvement is possible. The finding that no unit improved is therefore one-sided for every unit that began on the frontier and must be reported accordingly.

The second is the common shock. When all units are drawn from a single market observed over a single pair of periods, a change in trading conditions affecting the entire market moves the units together and shifts the frontier with them. A universal decline in measured efficiency is then consistent both with firm-specific investment decisions and with a market-wide compression of returns. The two cannot be distinguished without either a control group outside the affected market or a longer time series in which the pre-shock relationship is observable.

The third concerns inference. Bootstrap procedures for frontier estimators resample from the observed units to approximate the sampling distribution of the efficiency score [7]. With only six units, the resampling distribution reproduces the sample rather than the population, and confidence intervals constructed in this way would imply a precision that the data do not support. No interval estimates are reported here, and this omission is deliberate rather than an oversight.

Data and specifications

Panel construction

The panel comprises six platform-year observations: Alibaba Group for the fiscal years ending 31st March 2025 and 31st March 2026, and JD.com and PDD Holdings for the fiscal years ending 31st December 2024 and 31st December 2025. All figures are taken from the companies' own results announcements filed with the United States Securities and Exchange Commission and are reproduced in Table 1 at the precision reported. The reporting periods are not aligned: Alibaba's fiscal year ends on 31st March, meaning its FY2026 overlaps with the calendar years covered by the FY2025 observations of the other two platforms. This provides a further reason for interpreting cross-sectional comparisons with caution, though it does not affect the within-firm comparisons on which the paper's conclusions rest.

Pooling across years, rather than comparing three firms at a single point in time, is the principal method by which the analysis increases the number of observations. It treats a platform in a given year as a distinct decision-making unit, which is defensible when the technology and resource commitments differ materially between periods, as is the case here. This approach has a cost that should be acknowledged upfront rather than in the conclusion: Whether repeated observations of the same firm constitute genuinely independent units is a question the frontier literature has not settled, and the units in this panel are certainly not independent in the sense required for a sampling argument. Consequently, the subsequent scores are descriptive comparisons against a common benchmark, not estimates of a population parameter.

Input

A single input is used: research and development expenditure, reported by Alibaba Group as product development expenses. It is the one technology-investment measure disclosed on a comparable basis by all three firms in every period. Two limitations are attached to this measure. First, the underlying activity is not homogeneous across the three firms. Alibaba's product development expense spans a materially broader portfolio - encompassing cloud infrastructure, international commerce, and logistics alongside domestic marketplace operations - than the research and development lines of JD.com or PDD Holdings. Accordingly, this input measure is comparable in definition but not in scope. This constitutes a violation of the homogeneity requirement as defined by Cook et al. Together with the sample-related constraints noted earlier in the paper, it provides additional grounds to place interpretive emphasis on within-firm comparisons. Second, capital expenditure is excluded. It is economically significant for at least one platform: Alibaba reported quarterly capital expenditure of RMB 26,887 million for the quarter ended 31st March 2026 alone. However, it is not available on a comparable annual basis. Alibaba discloses it quarterly, JD.com reports a full-year figure (RMB 12,735 million in 2025 versus RMB 14,223 million in 2024), and PDD Holdings does not disclose it as a separate line item. Including it would require dropping two of the six units in the analysis. The direction of the resulting bias is known and is stated here explicitly rather than left to the reader's inference: Because the omitted input is largest at Alibaba, Alibaba's measured efficiency is overstated relative to the other two firms. Consequently, the cross-sectional ordering reported below is conservative with respect to that platform.

Outputs

Two outputs are used: revenue and net income attributable to ordinary shareholders. This selection is based on the rationale that a firm may convert investment into scale or into margin, and a specification that recognises only one would prejudge which conversion is more relevant. Net income attributable to ordinary shareholders, rather than a non-GAAP measure, is adopted as the baseline because it carries a uniform definition across all three firms. An earlier version of this analysis mixed bases, using non-GAAP net income for Alibaba and reported net income for the other two; this constituted a homogeneity violation on the output side and has been corrected here. Non-GAAP net income is retained and reported in Table 1, and the analysis has

been re-estimated using this metric to implement the first robustness check. During this correction, one figure was found to be misstated in the earlier version: Alibaba's non-GAAP net income for fiscal 2025 is RMB 158,122 million, as reported in the company's announcement of 15th May 2025, not the value previously used. All scores presented below are computed from the verified figures in Table 1.

Model choice and the units-to-dimensions ratio

With one input and two outputs, the specification has three dimensions against six units. This meets the Golany and Roll threshold of 2(m+s)=6 exactly and falls three units short of the Dyson et al. threshold of 3(m+s)=9. The shortfall is reported rather than concealed, and subsequent empirical testing investigates whether the estimator delivers meaningful discriminatory power in practice. The primary specification is an input-oriented constant-returns model, which is appropriate when the question is how much input could be saved while holding output constant. Consistent with Cook et al., this choice was fixed before the data were inspected. The variable-returns model is estimated and reported alongside it, both because the scale decomposition it permits is substantively informative and because the contrast between the two models itself provides evidence about the discriminating power of the panel.

Table 1. Inputs and outputs (RMB million).

Decision-making unit	R&D	Revenue	Net income attributable to ordinary shareholders	Non-GAAP net income
Alibaba FY2025	57,151	996,347	129,470	158,122
Alibaba FY2026	66,533	1,023,670	105,904	60,658
JD.com FY2024	17,031	1,158,819	41,359	47,800
JD.com FY2025	22,229	1,309,085	19,631	27,000
PDD Holdings FY2024	12,659	393,836	112,435	122,344
PDD Holdings FY2025	16,496	431,846	99,364	107,301

Sources: Alibaba Group (2025, 2026); JD.com (2026); PDD Holdings (2026). Alibaba's input is product development expenses. Fiscal years end 31 March for Alibaba and 31 December for JD.com and PDD Holdings. Baseline outputs are revenue and net income attributable to ordinary shareholders; non-GAAP net income is reported for application in the subsequent robustness-check analysis.

Results

The frontier, and whether the specification discriminates

Table 2 reports efficiency scores under both returns-to-scale assumptions, along with the reference set that determines each score. Under constant returns, two of the six units lie on the frontier, while four fall below it, with scores ranging from 0.299 to 1.000. This specification therefore provides meaningful discrimination, a real possibility enabled by the units-to-dimensions shortfall rather than a mere formality.

Under variable returns, it does not. Four of the six units are scored as efficient, and Alibaba FY2025 attains a score of unity despite being the second least efficient unit under constant returns. This is not a substantive finding about Alibaba; it is the extreme-point artifact described by Dyson et al. Alibaba FY2025 has the largest input in the panel, and under the convexity constraint, no combination of the remaining units can be constructed at a smaller scale to dominate it. Reporting the variable-returns column as if it measured pure technical efficiency at face value would credit Alibaba with an efficiency that the data do not establish. The column is reported here for precisely the opposite purpose: As evidence that, at this sample size, the variable-returns model is uninformative, and as the empirical justification for treating the constant-returns model as primary.

Table 2. Efficiency scores, scale decomposition and reference sets.

Decision-making unit	CRS θ	VRS θ	Scale efficiency	CRS reference set (λ)
Alibaba FY2025	0.371	1.000	0.371	JD.com FY2024 (0.535); PDD FY2024 (0.955)
Alibaba FY2026	0.299	0.684	0.437	JD.com FY2024 (0.644); PDD FY2024 (0.705)
JD.com FY2024	1.000	1.000	1.000	self (frontier)
JD.com FY2025	0.866	1.000	0.866	JD.com FY2024 (1.130)

Decision-making unit	CRS θ	VRS θ	Scale efficiency	CRS reference set (λ)
PDD Holdings FY2024	1.000	1.000	1.000	self (frontier)
PDD Holdings FY2025	0.740	0.781	0.948	PDD FY2024 (0.853); JD.com FY2024 (0.083)

Note: Input-oriented models, one input (R&D) and two outputs (revenue, net income attributable to ordinary shareholders). Scale efficiency is θ_CRS/θ_VRS. λ values are the optimal weights on peer units in the CRS programme. Computed by the author from the figures in Table 1.

The robust finding, and its mechanism

Efficiency fell at all three platforms between their two reported years: from 1.000 to 0.866 at JD.com, from 1.000 to 0.740 at PDD Holdings, and from 0.371 to 0.299 at Alibaba. No platform improved. Because business model, revenue recognition and accounting basis are held constant within each pair, the decline is not an artefact of comparison across heterogeneous firms.

The mechanism is transparent and can be stated without reference to the frontier at all. Table 3 reports the underlying growth rates. Research spending grew faster than revenue at every platform, by a wide margin in each case, and profit fell at every platform on both the reported and the non-GAAP basis. Any efficiency measure that treats research spending as an input and revenue and profit as outputs must register a decline under those conditions. The frontier analysis does not establish the direction of change - the growth rates do that on their own - but it quantifies the combined effect of the two output movements against a common benchmark and locates each platform relative to the best observed practice in the panel, which the growth rates cannot.

Table 3. Growth in inputs and outputs between the two reported years (%).

Platform (earlier year → later year)	R&D	Revenue	Net income	Non-GAAP net income
Alibaba (FY2025 → FY2026)	+16.4	+2.7	-18.2	-61.6
JD.com (FY2024 → FY2025)	+30.5	+13.0	-52.5	-43.5
PDD Holdings (FY2024 → FY2025)	+30.3	+9.7	-11.6	-12.3

Note: Computed by the author from the figures in Table 1.

Two features of the reference sets in Table 2 refine this picture. First, both frontier units are drawn from the earlier year: Every later-year unit is benchmarked against an FY2024 unit. Second, the reference set for JD.com FY2025 contains JD.com FY2024 alone, with λ=1.130, which is exactly the ratio of the two years' revenues. JD.com's score is therefore determined by the revenue constraint alone; profit plays no role in it. This is worth stating because the two-output specifications were justified by the argument that a firm may convert investment into scale or into margin. This rationale finds support in the evidence for PDD Holdings but carries no explanatory power whatsoever for JD.com.

What is not robust: The ranking of declines

Table 4 re-estimates the model under each defensible alternative. The direction of change is robust: Efficiency falls at every platform under the baseline, under the non-GAAP profit measure and under a revenue-only single-output model. Under variable returns it falls at Alibaba and PDD Holdings while JD.com is scored 1.000 in both years. This outcome stems from the censoring issue outlined earlier in the paper, instead of serving as evidence of stable performance.

The ranking of the declines is not robust. Under the baseline the steepest proportional decline belongs to PDD Holdings (-26.0% against -19.4% at Alibaba and -13.4% at JD.com); under the non-GAAP measure it belongs to Alibaba (-37.2%, against -26.4% and -13.4%); under the revenue-only model it belongs to PDD Holdings again (-15.9%), with Alibaba now showing the shallowest decline of the three (-11.7%). Alibaba moves from steepest to shallowest according to which of two equally defensible profit measures is adopted.

This is significant because an earlier version of this analysis presented the ranking as a substantive finding, noting that the depth of the decline correlated with the depth of each platform's infrastructural commitment. This correlation holds under the non-GAAP measure but not under the reported one. It is therefore retracted. What remains is the direction of the trend, which is consistent throughout, and the observation that Alibaba occupies the lowest position in both years under every specification that includes profit.

A third feature of Table 4 is worth noting. PDD Holdings'

frontier status depends entirely on the profit output: Under the revenue-only model, its score falls from 1.000 to 0.457 in FY2024 and from 0.740 to 0.385 in FY2025. PDD Holdings is efficient in terms of margin but markedly inefficient in terms of scale; JD.com is the reverse; Alibaba is interior on both counts. A single-output specification would not only have produced different numbers but would have led to a different interpretation of which platform was performing well and in what regard.

Table 4. Efficiency scores under alternative specifications.

Decision-making unit	(1) Baseline CRS	(2) Non-GAAP profit	(3) Revenue-only output	(4) VRS
Alibaba FY2025	0.371	0.389	0.256	1.000
Alibaba FY2026	0.299	0.244	0.226	0.684
JD.com FY2024	1.000	1.000	1.000	1.000
JD.com FY2025	0.866	0.866	0.866	1.000
PDD Holdings FY2024	1.000	1.000	0.457	1.000
PDD Holdings FY2025	0.740	0.736	0.385	0.781
Direction: efficiency fell at every platform?	Yes	Yes	Yes	Two of three
Ranking: steepest proportional decline	PDD (-26.0%)	Alibaba (-37.2%)	PDD (-15.9%)	Alibaba (-31.6%)

Specifications: (1) input-oriented CRS, outputs revenue and net income attributable to ordinary shareholders; (2) as (1) with non-GAAP net income substituted for reported net income; (3) as (1) with revenue as the sole output; (4) input-oriented VRS, outputs as in (1). Percentages in the final row are proportional changes in the efficiency score between the two reported years (computed by the author from the figures in Table 1).

Scale and pure technical efficiency

The scale decomposition in Table 2 provides a richer interpretation of Alibaba's position than a simple constant-returns score can. Alibaba's scale efficiency is 0.371 in FY2025 and 0.437 in FY2026. This reading suggests that almost all of its measured inefficiency stems from operating at a research expenditure scale far above the level at which the observed frontier technology converts it effectively, rather than from converting resources poorly at its own scale. In contrast, PDD Holdings has a scale efficiency of 0.948 in FY2025, so its decline is almost entirely attributable to a decline in pure technical efficiency.

This reading should be treated with caution for the reasons outlined earlier: Alibaba's variable-returns score of 1.000 in FY2025, which is the denominator of its scale efficiency, is partly an artefact of being an extreme point. The decomposition is offered as being consistent with a build-out account of Alibaba's position, not as a test of it. What it does establish is that the constant-returns score alone would support a stronger claim about managerial performance than the data warrant.

One structural qualification applies to the cross-sectional ordering throughout. JD.com's direct-sales model generates revenue that includes the value of goods purchased for resale, which inflates any revenue-based output measure relative to the marketplace revenue of the other two platforms. This is noted rather than corrected, as correcting it would require judgements that the method would then present as data. Its effect is visible in Table 4: JD.com's score is identical across all three specifications that vary the profit treatment because revenue is the binding constraint in each.

Discussion

Interpretation

If efficiency declines during capability building, a low score is not, in itself, evidence of poor management. The efficiency layer must therefore be read alongside the pathway layer rather than in isolation. A platform that is pursuing external provisioning should be expected to score low while provisioning is underway. The scale decomposition results give this expectation an empirical form, moving it from an assumption to an observation: A build-out represents a movement along the input axis and should therefore manifest as scale inefficiency paired with sound pure technical efficiency - a pattern that roughly matches Alibaba's observed results. The diagnostic question is whether efficiency scores rebound once capacity comes into full use. This question can only be answered through repeated deployment of the

efficiency analysis layer, and the concluding section outlines the requirements for carrying out such an exercise.

What cannot be identified here

The across-the-board efficiency decline documented in the empirical results has two potential explanations that cannot be disentangled under the present research design. The first is firm-specific: Each platform chose to increase research spending ahead of the revenue it would generate. The second is market-wide: The reported period was one in which competitive conditions in Chinese e-commerce compressed margins simultaneously for all three firms. The companies' own announcements support the second interpretation as much as the first. Alibaba attributes its decline in non-GAAP net income to "investment in technology businesses, quick commerce and user experiences" [8]. JD.com attributes its own decline to increased strategic investment in new initiatives [9]. PDD Holdings attributes its decline to channelling greater resources into business development [10]. Three firms describing the same competitive episode in similar terms is indicative of a common shock.

Separating the two explanations would require a research design this panel does not supply: a control group of comparable platforms outside the affected market segment, or a time series long enough that the pre-shock relationship between research spending and output is observable. In the absence of either, the responsible course is to report the decline, report that all three firms attribute it to competitive investment in the same period, and decline to attribute it to firm-specific management quality. The previously discarded ranking of efficiency decline magnitudes represented precisely this type of causal attribution, and its breakdown following a change in profit measurement metric serves as a reminder of the limited evidentiary support the data provide for such a claim.

A reporting protocol for the efficiency layer

The preceding sections yield a protocol for reporting this layer of the framework whenever it is applied in a concentrated market. Presenting a bare efficiency score in such a market lends the number an authority the data do not support; the following eight requirements are proposed as the minimum that keeps the limits visible.

(1) State the units-to-dimensions ratio against both the Golany-Roll and the Dyson thresholds, and report the shortfall where one exists.

(2) Estimate and report both constant- and variable-returns models. Where one of them fails to discriminate, report that failure as a result rather than selecting the model that yields the preferred number.

(3) Report the reference set and the optimal weights for every inefficient unit, so that a reader can see which comparisons drive each score.

(4) Identify which output constraint binds for each unit, and state where a multi-output specification does no work.

(5) Re-estimate the model under every defensible measurement choice - in particular every accounting basis available for an output - and report the whole set rather than one preferred specification.

(6) Distinguish explicitly between results that hold across the whole set and results that depend on a specification, and report the second class as specification-dependent rather than as findings.

(7) Where a unit lies on the frontier in the first period, note that its subsequent change is censored from above and that no inference about improvement is available for it.

(8) Where the units share a market and a period, state that a common shock cannot be separated from firm-specific effects, and refrain from causal attribution on that basis.

The protocol is restrictive by design. Its effect on the present paper is to reduce a set of six scores and a ranking to a single robust directional finding, a scale decomposition held loosely, and an explicit statement of what the data cannot support. That is a smaller result than the analysis initially appeared to yield, and a more defensible one.

Method choice

In markets containing a sufficient number of comparable units, frontier methods remain appropriate and powerful, and cross-sectional ranking is their natural output. Where such conditions are absent, the usual emphasis should be inverted: Within-unit longitudinal comparisons carry more weight than cross-unit rankings, and the analysis should be designed from the outset to prioritize the former. The present application illustrates why this is the case. Its cross-sectional ordering is compromised by three distinct issues: a non-homogeneous input measure, an omitted input that is largest at one platform, and a revenue measure that is not comparable across business

models. None of these issues affects the within-firm comparisons, each of which holds business model, accounting basis, and reporting scope constant. This inversion is the principal methodological lesson of this application, and it generalizes to any use of the PCP framework in a concentrated market.

Conclusion

This paper constructs the efficiency layer of the Pathway-Capability-Performance framework and applies it to a panel of six platform-years drawn from official filings. Two units lie on the constant-returns frontier. Efficiency declined at every platform between the two reported years, and this directional finding is robust to every alternative specification examined. The ranking of those declines, however, is not robust and is therefore reported as specification-dependent rather than as a substantive finding. A scale decomposition suggests that Alibaba's measured inefficiency is predominantly a matter of scale rather than of pure technical performance. This reading is offered as consistent with a build-out interpretation rather than as a direct test of it.

Seven limitations qualify these results. First, the panel is small, meeting one conventional units-to-dimensions threshold but falling three units short of the other. Second, the input measure is comparable in definition but not in scope, as Alibaba's product development expense spans a broader portfolio than the corresponding lines at the other two platforms. Third, capital expenditure is omitted because PDD Holdings does not disclose it separately, which likely overstates Alibaba's measured efficiency. Fourth, revenue is not adjusted for the distinction between direct sales and marketplace intermediation. Fifth, two reporting years cannot distinguish a build-out phase from a durable deterioration. Sixth, the units are not independent, and the frontier units are censored from above, so the finding that no platform improved is one-sided for two of the three. Finally, no interval estimates are reported, because bootstrap inference with only six units would convey a precision the data do not possess.

The most obvious extension would be a quarterly panel spanning several years. All three platforms report quarterly, and such a panel would raise the number of observations well above the relevant dimensionality thresholds. This, in turn, would make the variable-returns model informative rather than degenerate, permit a Malmquist decomposition of productivity change into efficiency-change and frontier-shift components, and render bootstrap inference defensible. A longer time series would also capture the pre-shock relationship between research spending and output. This single supplementary piece of information would make it possible to disentangle explanations stemming from market-wide common shocks from those driven by firm-specific factors. The independence question would also have to be addressed explicitly in such an extension, rather than merely noted as a limitation on how the efficiency scores may be interpreted.

Funding

This work was not supported by any funds.

Acknowledgements

The author would like to show sincere thanks to those technicians who have contributed to this research.

Conflicts of Interest

The author declares no conflict of interest.

References

[1] Verhoef, P. C., Broekhuizen, T., Bart, Y., Bhattacharya, A., Dong, J. Q., Fabian, N., Haenlein, M. (2021) Digital transformation: a multidisciplinary reflection and research agenda. *Journal of Business Research*, 122, 889-901.

[2] Charnes, A., Cooper, W. W., Rhodes, E. (1978) Measuring the efficiency of decision making units. *European Journal of Operational Research*, 2(6), 429-444.

[3] Banker, R. D., Charnes, A., Cooper, W. W. (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*, 30(9), 1078-1092.

[4] Golany, B., Roll, Y. (1989) An application procedure for DEA. *Omega*, 17(3), 237-250.

[5] Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., Shale, E. A. (2001) Pitfalls and protocols in DEA. *European Journal of Operational Research*, 132(2), 245-259.

[6] Cook, W. D., Tone, K., Zhu, J. (2014) Data envelopment analysis: prior to choosing a model. *Omega*, 44, 1-4.

[7] Simar, L., Wilson, P. W. (1998) Sensitivity analysis of efficiency scores: How to bootstrap in

nonparametric frontier models. *Management Science*, 44(1), 49-61.

[8] Rodchenkov, M. V. (2026) The functionality of economic metrics in a multipolar world. *Journal of New Economy*, 27(1), 115-133.

[9] Abdallah, T., Bray, R. L., Choi, H., Glinskiy, V. (2026) The impact of market growth on delivery speed: evidence from JD. com. *Service Science*, 18(2), 96-119.

[10] Reust, D. (2026) Marketing strategies of Chinese online platforms in Switzerland. *Innovating Business and Education for Sustainable Development,* 173-187.