Contemporary Mathematics
https://ojs.wiserpub.com/index.php/CM/

Research Article

DEA-Inspired Constrained Log-Log Quantile Frontier: Smooth Benchmarking, Calibration, and Dynamic Interpretation

Vladislav Spitsin[1*], Nikita Martyushev[2], Lubov Spitsina[1], Magerram Gasanov[1], Victoria Leonova[1]

[1] Business School, Tomsk Polytechnic University, Tomsk, Russia
[2] Engineering School of Information Technology and Robotics, Tomsk Polytechnic University, Tomsk, Russia
E-mail: spicynvv@tpu.ru

Received: 21 November 2025; **Revised:** 2 February 2026; **Accepted:** 9 February 2026

Abstract: This paper proposes a Data Envelopment Analysis (DEA)-inspired smooth benchmarking approach based on a constrained log-log quantile production frontier. The frontier is estimated by quantile regression under economically motivated inequality restrictions—monotonicity in inputs and a non-increasing-returns (concavity-compatible) restriction within the Cobb-Douglas class—yielding a continuously differentiable benchmark with elasticity-based interpretation and tractable inference via constrained quantile regression theory. Our contribution is threefold: we (i) formalize large-sample inference for constrained quantile frontiers under economically interpretable inequality restrictions, (ii) introduce a calibration perspective for probabilistic frontiers via exceedance/coverage diagnostics (and a simple non-crossing adjustment for multiple quantiles), and (iii) derive closed-form links between output- and input-oriented quantile efficiency indices through the returns-to-scale parameter, together with a high-quantile interpretation within a one-sided stochastic production model. We emphasize the probabilistic nature of quantile frontiers: for any fixed quantile level $\tau \in (0, 1)$, the fitted frontier is a coverage benchmark rather than a deterministic envelopment surface, so a non-negligible share of observations may lie above it. Empirically, we apply the framework to firm-level panel data from the SPARK-Interfax information system covering 1,035 Russian manufacturing firms over 2019–2023 (5,175 firm-year observations). We benchmark the proposed smooth quantile frontier against classical DEA and a parametric Stochastic Frontier Analysis (SFA), and we report internal validation (pinball loss, coverage) together with sensitivity diagnostics across τ. The resulting efficiency measures exhibit significant associations with profitability (net Return on Assets (ROA)), changes in profitability, and sales growth, indicating economic relevance. Overall, the proposed approach complements deterministic envelopment by providing smooth differentiability, robustness to noise, and calibration-driven interpretability for heterogeneous datasets, while retaining economically meaningful shape discipline and enabling a dynamic decomposition of frontier shifts and relative performance.

***Keywords*:** constrained quantile regression, log-log quantile frontier, Data Envelopment Analysis (DEA) benchmarking, shape restrictions, calibration, non-crossing quantiles, technical efficiency, dynamic decomposition

MSC: 62P20, 62J20, 90C31, 90B30, 62F15

DOI: https://doi.org/10.37256/cm.7320269045

1. Introduction

1.1 *Background and motivation*

Measuring production efficiency and identifying the technological frontier remain central topics in empirical productivity analysis. Traditional approaches—such as Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA)—have provided powerful tools for evaluating firm performance but rely on strong assumptions about functional form and data regularity. In many real-world datasets, particularly those including firms of very different sizes, these assumptions lead to instability and lack of smoothness in estimated frontiers.

Recent advances in quantile regression and semi-parametric modeling offer a promising alternative. By estimating conditional quantiles of output, these methods capture heterogeneous production behavior and account for stochastic noise without imposing rigid parametric structures. However, the relationship between deterministic DEA envelopes and stochastic quantile frontiers remains insufficiently formalized. Rather than seeking a strict equivalence, a useful goal is to develop a smooth, probabilistic benchmarking frontier that is economically shape-consistent with DEA while remaining statistically tractable. This DEA-inspired perspective provides both theoretical and practical advantages—offering a smooth and data-driven framework for analyzing firm-level efficiency and frontier dynamics under heterogeneity and noise.

1.2 *Literature review*

The measurement of Technical Efficiency (TE) and estimation of the production frontier are long-standing topics in operations research, econometrics, and mathematical optimization. Two paradigms dominate this literature: the deterministic nonparametric approach, represented by DEA, and the stochastic parametric approach, represented by SFA. Both attempts to approximate the boundary of feasible production, yet differ in their treatment of noise, convexity, and inference [1–3].

The deterministic framework, initiated by Charnes et al. [2] and extended to Variable Returns to Scale (VRS) by Banker et al. [3], uses linear programming to envelop observed data points. Over the past decades DEA has evolved into a vast family of models including input/output orientation, bootstrap inference, robust order-m frontiers, and directional measures [4–7]. Recent works highlight DEA's flexibility for public-sector and healthcare efficiency [8] and methodological innovations such as ranking-based second-stage estimation [9]. Nevertheless, classical DEA remains deterministic and cannot directly accommodate random noise. A related line of research develops network and multi-stage DEA models that open the "black box" of production by explicitly modelling intermediate products and sub-technologies (network DEA), thereby enabling efficiency analysis in multi-process systems and dynamic settings [10].

The stochastic paradigm, initiated by Aigner et al. [1], interprets inefficiency as part of a composed error term. Modern SFA research incorporates Bayesian, panel, and quantile-based extensions [11–17]. Alongside classical parametric SFA, stochastic semi-/nonparametric frontier methods have been proposed to combine axiomatic shape restrictions (e.g., monotonicity/concavity) with stochastic noise, often formulated via convex regression and related mathematical programming tools (e.g., Stochastic Nonparametric Envelopment of Data (StoNED)/Convex Nonparametric Least Squares (CNLS)-type approaches) [18, 19].

Quantile-based formulations have become increasingly important because they characterize heterogeneous produc-tion behavior and are robust to outliers [20–23]. Recent studies have increasingly applied robust and quantile-based estimation methods to improve the stability of statistical models under heterogeneous data distributions [24]. Quantile regression belongs to the family of statistical learning methods, where the conditional quantile function is estimated through convex optimization under an asymmetric loss [20, 25]. In this sense, quantile frontier estimation can be interpreted as a data-driven learning approach to the production boundary, conceptually related to machine-learning frameworks that infer functional relationships directly from data. This view aligns quantile frontiers with modern developments in statistical learning theory, emphasizing their flexibility and generalization capacity relative to classical deterministic envelopes.

Machine-learning frontier models. In parallel, frontier analysis has increasingly incorporated modern Machine-Learning (ML) tools to capture nonlinearities, high-dimensional covariates, and complex heterogeneity. Recent work includes ML-driven stochastic frontier modelling (e.g., neural-network-based approaches within SFA-type objectives) and hybrid DEA-ML pipelines used for prediction or profiling of efficiency patterns. These methods can improve predictive performance but typically require careful regularization and may weaken interpretability unless combined with explicit economic structure or shape constraints [26, 27].

Robustness and diagnostics. Because frontier estimators can be sensitive to outliers and sampling variability, the literature has developed robustness and diagnostic techniques, including partial-frontier ideas (order-m and α-frontiers), trimming/influence diagnostics, and bootstrap-based inference and sensitivity analysis for DEA-type models [28, 29]. In probabilistic frontier settings, coverage/exceedance diagnostics and (when multiple quantiles are estimated) non-crossing adjustments provide natural calibration and robustness checks complementary to deterministic diagnostics.

Table 1 summarizes the main strands of research on frontier and efficiency estimation, highlighting the conceptual differences between deterministic, stochastic, quantile-based, and ML-augmented/multi-stage approaches.

Table 1. Core strands in frontier and efficiency estimation

Strand	Conceptual idea	Representative references
DEA (deterministic)	Linear-programming frontier; input/output orientation; Constant Returns to Scale (CRS)/VRS returns; no noise	Charnes et al. [2]; Banker et al. [3]; Zarrin and Brunner [6]; Mergoni et al. [7]
Robust/partial DEA	Order-m or α-frontile ideas; robustness to outliers	Cazals et al. [4]; Daraio and Simar [5]; Lv et al. [30]
SFA (parametric)	Composed-error stochastic frontier; parametric inefficiency distribution	Aigner et al. [1]; Tsionas and Kumbhakar [11]; Fusco et al. [13]
Quantile regression (general)	Pinball-loss minimization; conditional quantiles as flexible functions	Koenker and Bassett [20]; Chernozhukov et al. [25]
Quantile frontiers/SFA (semi-parametric)	Upper-tail quantiles as stochastic frontiers; shape constraints	Bondell et al. [31]; Jiang and Yu [32]; Jradi et al. [21]; Tsionas [12]; Dai and Kuosmanen [23]; Lee et al. [16]; Wei et al. [17]
Network/multi-stage DEA	Internal structure (sub-technologies, intermediate products), dynamic links across stages	Färe et al. [10]
Stochastic semi-/nonparametric envelopment (convex regression or StoNED-type)	Shape-restricted nonparametric frontiers with stochastic noise via mathematical programming	Kuosmanen [18]; Kuosmanen and Johnson [19]; Dai et al. (pyStoNED) [33]
Diagnostics/bootstrap inference (DEA-type)	Sensitivity analysis and valid inference for nonparametric efficiency estimators; two-stage pitfalls and fixes	Simar and Wilson [28]; Simar and Wilson [29]
ML-assisted frontier modelling	Flexible learners (Neural Networks/Random Forest (NN/RF)) combined with DEA/SFA objectives; prediction and heterogeneity modelling	Tsionas [26]; Chen et al. [27]

Source: Compiled by the authors based on research materials

Table 1 illustrates that efficiency research now spans deterministic, stochastic, and hybrid approaches. Among the many approaches to efficiency measurement, DEA remains the dominant nonparametric framework [7], while quantile-based regression frontiers have recently emerged as a probabilistic generalization bridging DEA and stochastic frontier analysis [14]. DEA provides a convex-envelope representation of technology with explicit monotonicity and VRS shape constraints [3, 5]. Quantile regression, by contrast, describes the conditional distribution of output given inputs through asymmetric loss minimization [20, 25]. Both share a geometric interpretation as boundaries of conditional distributions but differ in inference and noise handling. Recent advances aim to link the two paradigms via shape-restricted and non-crossing quantile estimators [22, 23, 31, 32] or by embedding quantile structures into stochastic frontiers [12, 13, 15]. In this paper we adopt a DEA-inspired viewpoint: we impose economically meaningful shape restrictions in a log-log quantile frontier to obtain a smooth benchmarking boundary that is compatible with core DEA axioms (such as monotonicity and technology convexity through concavity conditions within the chosen parametric class), while explicitly treating the frontier as probabilistic rather than as a deterministic envelope.

Beyond core theory, frontier methods are widely applied across sectors (e.g., banking, manufacturing, energy, transport, health care, and environment), and modern computational toolkits have facilitated broader empirical uptake of shape-constrained and stochastic semi-/nonparametric frontiers [18, 33].

Building on this overview, Table 2 reviews recent studies focusing on quantile-based frontier estimation, emphasizing the shift toward stochastic and log-log formulations, and situates them relative to ML-based and multi-output or network frontier developments that motivate extensions of our approach.

Table 2. Recent research on quantile-based frontier estimation

Study	Model specification	Frontier type	Main contribution	Relation to this paper
Koenker and Bassett [20]	Linear quantile regression	General conditional quantiles	Introduced quantile regression via asymmetric (pinball) loss	Foundational concept; no frontier interpretation
Bondell et al. [31]	Quantile regression with non-crossing constraints	Semi-parametric	Ensures monotonic quantile curves	Basis for multi-quantile shape-restricted benchmarking
Lee et al. [16]	Quantile stochastic frontier (parametric, log-log)	Stochastic	Bayesian estimation; allows endogeneity	Uses log-log form, but no explicit DEA-inspired shape restrictions
Jradi et al. [21]	Quantile SFA with normal-exponential inefficiency	Stochastic	Estimates upper-tail production frontiers	Parametric version without shape restriction
Papadopoulos and Parmeter [14]	Quantile SFA survey	Stochastic	Review of quantile frontiers and estimation issues	Highlights heterogeneity; positions quantiles as stochastic benchmarks
Fusco et al. [13]	Parametric quantile frontier	Stochastic	Joint modelling of quantile coefficients	Focus on estimation; limited emphasis on calibration diagnostics
Dai and Kuosmanen [23]	Shape-constrained quantile functions	Semi-parametric	Generalized quantile/expectile properties	Provides tools for monotonicity; no efficiency benchmarking focus
Chernozhukov et al. [25]	Log-log quantile regression (general)	Non-frontier	Shape properties under regularity conditions	Basis for our assumptions and log-log interpretation
Zheng et al. [15]	ML-based nonparametric quantile frontier	Stochastic/ML	Flexible quantile estimation via learning algorithms	Motivates ML extensions; does not impose DEA-inspired constraints
Kuosmanen [18]	StoNED (stochastic semi-/nonparametric), applications	Stochastic nonparametric	Shape restrictions with noise; regulatory or sector applications	Complementary frontier family; motivates broader benchmarking comparisons
Kuosmanen and Johnson [19]	Multi-output StoNED (directional distance)	Stochastic multi-output	Joint production or directional distance under shape restrictions	Motivates multi-output/network extensions beyond single-output case
Dai et al. [33]	Software for convex regression/StoNED/convex quantile regression	Computational toolkit	Open-access implementation enabling applied work across sectors	Supports scalability and practical implementation discussion
Tsionas [26]	Bayesian Artificial Neural Network (ANN) frontier model	ML/stochastic	Flexible ML frontier bridging DEA/SFA ideas	Benchmarks the "ML frontier" direction; contrasts interpretability vs flexibility
This paper	Constrained log-log quantile frontier (DEA-inspired shape restrictions)	Stochastic benchmarking and DEA comparison	Constrained estimation, calibration/coverage diagnostics, smooth benchmarking, and dynamic interpretation	Complements DEA by providing a smooth probabilistic benchmark and empirical validation

Source: Compiled by the authors based on research materials

The growing use of quantile-based frontiers reflects the convergence of econometric, optimization, and machine-learning perspectives on efficiency measurement. Most recent studies [12, 13] adopt log-log or semi-parametric forms but focus primarily on estimation and robustness rather than explicit calibration and benchmarking protocols. While [25] established general properties of log-log quantile functions such as monotonicity and shape properties under regularity conditions, these results were not derived in an efficiency-analysis context. Our paper builds directly on this theoretical foundation, introducing a constrained log-log quantile frontier that incorporates economically interpretable shape restrictions and provides a smooth stochastic benchmarking boundary inspired by DEA. We formalize the probabilistic (coverage) interpretation of quantile frontiers, provide calibration diagnostics, and position the approach as a statistically tractable complement to deterministic envelopment methods.

1.3 *Research problem and contributions*

Research gap. Despite progress, two major issues persist:

• Theoretical bridge. Quantile frontier models rarely impose DEA-consistent shape restrictions (monotonicity, convexity, VRS). A rigorous log-log quantile specification with economically interpretable inequality restrictions (e.g., monotonicity and non-increasing-returns/concavity conditions within a parametric class) and with transparent probabilistic interpretation remains underdeveloped for practical benchmarking.

• Validation and interpretability. Comparative assessments between deterministic and stochastic frontiers seldom link mathematical efficiency with real-world outcomes such as profitability or sales growth. Systematic validation frameworks combining internal metrics (pinball loss, coverage) and external criteria (economic performance) are missing.

This paper contributes in three ways:

• Model formulation. It proposes a DEA-inspired constrained log-log quantile frontier that enforces economically meaningful shape restrictions (monotonicity and non-increasing-returns/concavity within the Cobb-Douglas class) and yields a smooth benchmarking boundary with elasticity-based interpretation.

• Mathematical theory. We establish identifiability and the correct large-sample behavior of the constrained quantile regression estimator under inequality restrictions, including the role of active constraints. We also provide a high-quantile interpretation within a one-sided stochastic production model, supporting the use of large quantiles as near-frontier benchmarks within the model class.

• Model validation. We evaluate models via internal predictive accuracy and coverage/exceedance calibration and external economic relevance (ROA, ΔROA, sales growth), linking statistical efficiency to economic outcomes.

From an economic standpoint, the proposed framework addresses a recurrent limitation of traditional DEA: its deterministic treatment of efficiency under potentially noisy or heterogeneous environments. In empirical production data—especially at the firm or sectoral level—inputs and outputs often exhibit high dispersion, measurement error, and structural heterogeneity. Under such conditions, DEA's piecewise-linear envelope may either overfit outliers or underestimate efficiency for small or volatile units. By contrast, the log-log quantile frontier provides a stochastic, smooth, and scale-consistent representation of technology, capable of distinguishing genuine inefficiency from random shocks. The logarithmic specification ensures elasticity-based interpretation of coefficients, while quantile estimation naturally accounts for heterogeneity across producers.

Mathematically, this formulation maintains key economic shape restrictions aligned with DEA intuition—notably monotonicity and concavity-compatible behavior within the chosen parametric class—yet introduces continuous differentiability and statistical inference absent in deterministic models. Consequently, the log-log quantile frontier retains DEA-style benchmarking interpretability but enhances robustness to data irregularities and small-sample noise. It can thus be viewed as a DEA-inspired probabilistic benchmark: a smooth frontier that is calibrated by conditional coverage and remains well-behaved when observations are dispersed or uncertain.

Economically, this duality offers several practical advantages:

• Noise tolerance. Random fluctuations in output or inputs are absorbed through the quantile structure rather than producing extreme frontier shifts.

• Heterogeneity modeling. Firms operating under different technologies are represented through different conditional quantiles.

• Elasticity interpretation. Log-log coefficients yield direct scale and substitution elasticities, enabling comparison across technologies and time.

• Empirical validation. Efficiency estimates can be statistically linked to profitability, growth, or innovation indicators—bridging mathematical efficiency with observable performance.

Overall, the paper contributes to both mathematical and applied efficiency analysis by demonstrating that a log-log quantile specification:

• Provides a smooth probabilistic benchmarking frontier inspired by DEA, with monotonicity and concavity-compatible shape restrictions within the Cobb-Douglas class;

• Supports calibration and model checking via coverage/exceedance diagnostics (and non-crossing adjustment when multiple quantiles are used);

• Enables both input- and output-oriented efficiency benchmarking with a transparent mapping between orientations in the parametric model;

• Adds stochastic flexibility and smoothness required for inference on real economic data with wide variability.

Thus, the model serves as a mathematically grounded and economically interpretable DEA-inspired smooth benchmarking framework, positioned at the intersection of efficiency analysis and modern statistical learning.

2. Theoretical framework

This section introduces (i) the deterministic DEA technology and efficiency concepts, (ii) a stochastic production model and conditional-quantile frontiers, (iii) a log-log (Cobb-Douglas) quantile frontier with economically meaningful shape restrictions, (iv) efficiency scores induced by a quantile frontier (with a correct treatment of observations above the frontier), and (v) a simple dynamic extension. Notation follows standard efficiency-analysis conventions [5, 34, 35].

2.1 *Production technology and deterministic efficiency*

Let $X = (X_1, \ldots, X_m)^\top \in \mathbb{R}_+^m$ denote inputs and $Y \in \mathbb{R}_+$ the (single) output. The production technology is represented by a set

$$\mathscr{T} \subseteq \mathbb{R}_+^m \times \mathbb{R}_+, \tag{1}$$

interpreted as all feasible input-output combinations (x, y).

The (population) deterministic frontier (output function) is defined by

$$f_0(x) = \sup\{y \geq 0 : (x, y) \in \mathscr{T}\}, \quad x \in \mathbb{R}_+^m. \tag{2}$$

A common maintained economic structure is free disposability (if $(x, y) \in \mathscr{T}$, then $(x', y') \in \mathscr{T}$ for all $x' \geq x$, $0 \leq y' \leq y$), implying that f_0 is coordinate-wise nondecreasing. In many applications (and in standard DEA constructions) the technology is also convex, which implies that the boundary f_0 is concave as a function of inputs.

For an observation (x_i, y_i), the input-oriented Farrell efficiency (radial contraction) is

$$\mathrm{TE}_i^I = \inf\{\theta \in (0, 1] : (\theta x_i, y_i) \in \mathscr{T}\}. \tag{3}$$

The output-oriented Farrell efficiency (radial expansion) is

$$\mathrm{TE}_i^O = \sup\{\varphi > 0 : (x_i, \varphi y_i) \in \mathscr{T}\}. \tag{4}$$

In deterministic DEA, $\mathscr{T}$ and f_0 are estimated by piecewise-linear envelopment under maintained returns-to-scale assumptions (CRS/VRS) [2, 3, 34]. Importantly, the DEA frontier is an envelope of the sample; in contrast, a quantile frontier introduced below is not an envelope unless $\tau \to 1$ and additional structure is imposed.

2.2 *Stochastic production and conditional-quantile frontiers*

To accommodate noise, heterogeneity, and unobserved shocks, we treat Y as random conditional on X. A convenient specification is multiplicative in levels and additive in logs [12, 13, 20, 21]:

$$\log Y = q_\tau(X) + U_\tau, \tag{5}$$

where $q_\tau(X)$ is the conditional τ-quantile of $\log Y$ given X, and U_τ is a residual with

$$Q_\tau(U_\tau \mid X) = 0. \tag{6}$$

Equivalently, defining the conditional τ-quantile of $\log Y$ as

$$Q_\tau(\log Y \mid X = x) = q_\tau(x), \quad \tau \in (0,\, 1), \tag{7}$$

the associated τ-quantile frontier in levels is

$$f_\tau(x) = \exp\{q_\tau(x)\}. \tag{8}$$

A key interpretation is conditional coverage:

$$\mathbb{P}(Y \le f_\tau(X) \mid X) = \tau \text{ (under correct specification and mild regularity).} \tag{9}$$

Hence, for any fixed $\tau < 1$, the quantile frontier does not generally bound the data from above; a nontrivial share $(1-\tau)$ of observations is expected to lie above it.

2.3 *Log-log (Cobb-Douglas) quantile frontier with shape restrictions*

We adopt a log-log parametric representation of the conditional quantile:

$$q_\tau(x) = \beta_0(\tau) + \sum_{j=1}^{m} \beta_j(\tau)\, \log x_j, \quad x \in \mathbb{R}_+^m, \tag{10}$$

which yields the Cobb-Douglas quantile frontier in levels:

$$f_\tau(x) = \exp\{\beta_0(\tau)\} \prod_{j=1}^{m} x_j^{\beta_j(\tau)}. \tag{11}$$

We impose economically meaningful shape restrictions:
• Monotonicity (free disposability in inputs):

$$\beta_j(\tau) \geq 0, \quad j = 1, \ldots, m. \tag{12}$$

• Non-Increasing Returns to Scale (NIRS)/sub-homogeneity and concavity (sufficient condition):

$$\kappa(\tau) \equiv \sum_{j=1}^{m} \beta_j(\tau) \leq 1. \tag{13}$$

Under (12)–(13), f_τ is coordinate-wise nondecreasing. Moreover, for Cobb-Douglas, (12)–(13) is a standard sufficient condition for concavity on $\mathbb{R}_+^m$, making the frontier compatible with convex technology sets (in the sense of concave boundary functions). The classification by $\kappa(\tau)$ is classical within Cobb-Douglas:
• $\kappa(\tau) = 1$: Constant Returns to Scale (CRS);
• $\kappa(\tau) < 1$: Decreasing/Non-Increasing Returns to Scale (DRS/NIRS);
• $\kappa(\tau) > 1$: Increasing Returns to Scale (IRS), not allowed if we enforce (13).

Figure 1 illustrates, using simulated data, the geometric contrast between the piecewise-linear DEA envelope and a smooth log-log quantile frontier in a single-input, single-output setting.

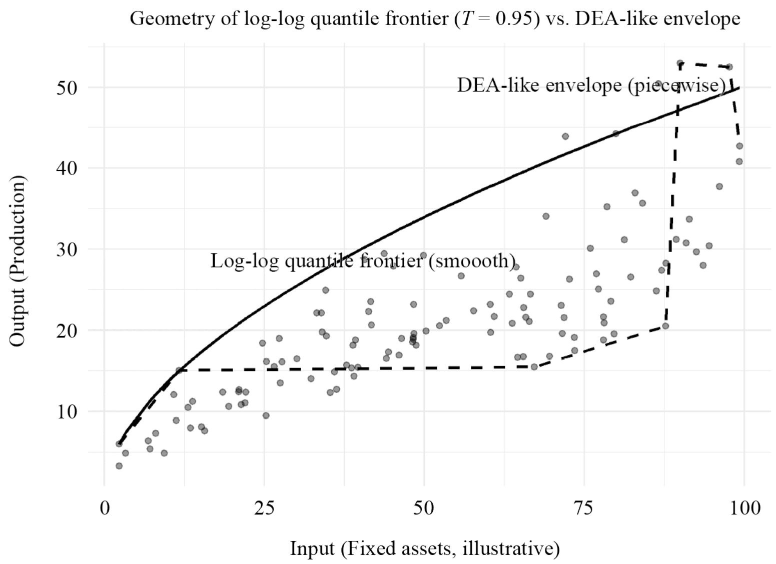

Figure 1. Geometry of the log-log quantile frontier and the DEA envelope (simulated data)
Source: Calculated by the authors

The solid curve represents the estimated $\tau = 0.95$ log-log quantile frontier, while the dashed line depicts the piecewise-linear DEA envelope computed from the same observations. The quantile frontier provides a smooth, differentiable probabilistic benchmarking curve that can be estimated by convex optimization and can incorporate monotonicity and the NIRS/concavity restriction $\kappa(\tau) \leq 1$, whereas the DEA envelope serves as a deterministic envelopment reference.

The log-log formulation implies continuous differentiability of the fitted frontier: unlike the piecewise-linear DEA envelope, the quantile frontier defines a smooth surface with well-behaved gradients, which supports sensitivity analysis and gradient-based interpretation. In addition, the log-log specification yields elasticity-based quantities (returns-to-scale and input elasticities) that are central for interpreting production technology and efficiency. In Section 4, analogous figures demonstrate these relationships on the empirical data.

Remark (VRS versus NIRS).

Constraint $\kappa(\tau) \leq 1$ should be interpreted as NIRS/concavity within this parametric class. It is not equivalent to DEA-VRS (which is defined by a different convexity/affine restriction in envelopment models). Our parametric frontier may approximate a VRS-type boundary empirically, but the notions should not be conflated.

2.4 *Estimation via constrained quantile regression in log-log form*

Let $Z = (1, \log X_1, \ldots, \log X_m)^\top \in \mathbb{R}^{m+1}$ and $W = \log Y$. For a fixed $\tau \in (0, 1)$, we estimate $\beta(\tau) = (\beta_0(\tau), \ldots, \beta_m(\tau))^\top$ by constrained quantile regression:

$$\widehat{\beta}(\tau) \in \arg\min_{\beta \in \mathscr{B}} \sum_{i=1}^{n} \rho_\tau \left(W_i - Z_i^\top \beta\right), \tag{14}$$

where $\rho_\tau(u) = u(\tau - \mathbf{1}\{u < 0\})$ is the check loss and the feasible set is

$$\mathscr{B} = \left\{\beta \in \mathbb{R}^{m+1} : \beta_j \geq 0 \ (j = 1, \ldots, m), \quad \sum_{j=1}^{m} \beta_j \leq 1\right\}. \tag{15}$$

Then the estimated τ-frontier is

$$\widehat{f}_\tau(x) = \exp\{\widehat{\beta}_0(\tau)\} \prod_{j=1}^{m} x_j^{\widehat{\beta}_j(\tau)}. \tag{16}$$

2.5 *Efficiency scores induced by a quantile frontier (correct handling of "above-frontier" points)*

Because a quantile frontier is not an envelope for $\tau < 1$, a ratio $y_i/\widehat{f}_\tau(x_i)$ can exceed 1. We therefore distinguish two objects:

1. Exceedance ratio (unbounded):

$$R_{i,\tau} = \frac{y_i}{\widehat{f}_\tau(x_i)}. \tag{17}$$

Large $R_{i,\tau} > 1$ indicates that observation i lies above the estimated τ-frontier. The fraction of such exceedances is informative for calibration.

2. Bounded output-oriented "technical efficiency score":

$$\mathrm{TE}_{i,\tau}^{O} = \min\left\{1, \frac{y_i}{\widehat{f}_\tau(x_i)}\right\}. \tag{18}$$

This produces a DEA-like bounded index in $(0, 1]$, while remaining consistent with the probabilistic nature of quantile frontiers.

For input orientation, we consider radial scaling of inputs $x \mapsto \theta x$ (with $\theta \in (0, 1]$) and define the bounded input-oriented score

$$\mathrm{TE}^{I}_{i,\,\tau} = \inf\{\theta \in (0, 1] : y_i \le \widehat{f}_\tau(\theta x_i)\}. \tag{19}$$

For Cobb-Douglas frontiers this admits a closed form (see Proposition 5 in Section 3).

2.6 *Interpretation of τ and relation to partial frontiers*

Parameter τ has a clear probabilistic meaning: it is the conditional coverage level in (9). In frontier terms, increasing τ moves the frontier upward and reduces the expected exceedance probability.

Important caution. Partial (order-α) frontiers and τ-quantile frontiers are related conceptually (both provide robust, non-extreme frontiers), but they are not generally equivalent objects. In particular, one should not claim a universal one-to-one correspondence "order-α = quantile-τ" without additional restrictive assumptions. In this paper we therefore treat τ as a coverage/robustness parameter rather than as an exact proxy for order-α.

2.7 *Dynamic quantile frontier and intertemporal decomposition*

For panel or repeated cross-sections indexed by $t \in \{1, \ldots, T\}$, we allow coefficients to vary over time:

$$q_{\tau,\,t}(x) = \beta_{0,\,t}(\tau) + \sum_{j=1}^{m} \beta_{j,\,t}(\tau)\log x_j, \quad f_{\tau,\,t}(x) = \exp\{\beta_{0,\,t}(\tau)\}\prod_{j=1}^{m} x_j^{\beta_{j,\,t}(\tau)}. \tag{20}$$

For a given unit i, a convenient multiplicative decomposition of output change relative to the τ-frontier is

$$\frac{y_{i,\,t+1}}{y_{i,\,t}} = \underbrace{\frac{f_{\tau,\,t+1}(x_{i,\,t+1})}{f_{\tau,\,t}(x_{i,\,t})}}_{\text{frontier shift + input mix/scale effect}} \times \underbrace{\frac{R_{i,\,\tau,\,t+1}}{R_{i,\,\tau,\,t}}}_{\text{relative position change}}, \tag{21}$$

where $R_{i,\,\tau,\,t} = y_{i,\,t}/f_{\tau,\,t}(x_{i,\,t})$. When bounded scores are desired, $R_{i,\,\tau,\,t}$ can be replaced by $\mathrm{TE}^{O}_{i,\,\tau,\,t}$ at the cost of censoring exceedances.

Figure 2 demonstrates, using generated data, the concept of frontier dynamics and intertemporal efficiency adjustments between consecutive periods.

The simulated frontiers $f_t(x)$ and $f_{t+1}(x)$ illustrate Technical Change (TC) through an outward shift of the benchmark frontier and Efficiency Change (EC) through firms' movement relative to the new benchmark. The figure is intended as a conceptual illustration of the dynamic decomposition used later in the empirical section (Section 4) on observed firm data.

The next section develops formal mathematical results for the constrained log-log quantile frontier—including identifiability and asymptotic properties within the chosen model class—and highlights how the proposed specification supports smooth stochastic benchmarking under economically interpretable shape restrictions. In particular, the log-log form yields a differentiable frontier and a transparent scaling index that facilitate dynamic interpretation and sensitivity analysis.

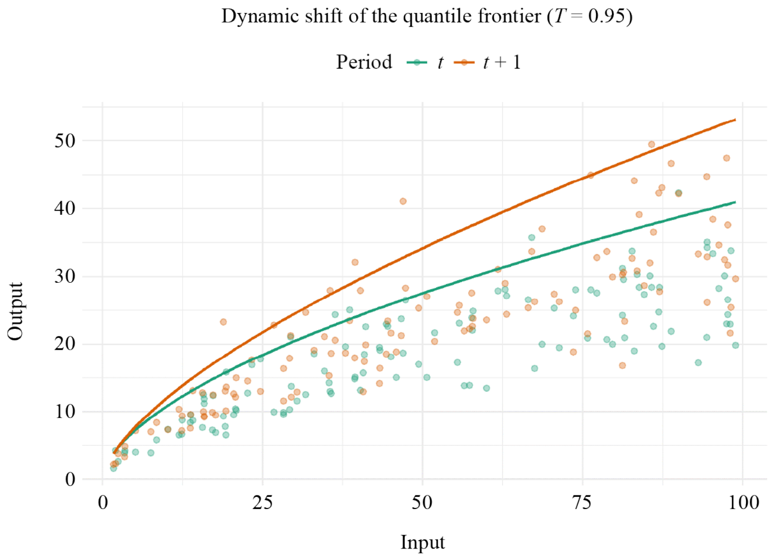

Figure 2. Dynamic shift of the log-log quantile frontier (simulated data)
Source: Compiled by the authors based on research materials

3. Mathematical results

This section states formal results underpinning the constrained quantile frontier and provides several useful closed-form expressions.

3.1 *Assumptions*

Fix $\tau \in (0, 1)$. Let $(W_i, Z_i)_{i=1}^n$ be i.i.d., where $W_i = \log Y_i$ and $Z_i = (1, \log X_{i1}, \dots, \log X_{im})^\top$.

A1 (Regularity of design). $\mathbb{E} \parallel Z \parallel^2 < \infty$ and $\mathbb{E}[ZZ^\top]$ is positive definite.

A2 (Well-defined conditional quantile). The conditional density of W given Z exists and is continuous in a neighborhood of $Z^\top \beta(\tau)$, with $f_{W|Z}(Z^\top \beta(\tau) \mid Z)$ bounded away from 0 and ∞ almost surely.

A3 (Correct specification at τ). $Q_\tau(W \mid Z) = Z^\top \beta(\tau)$ for some $\beta(\tau) \in \mathscr{B}$.

A4 (Feasible set). $\mathscr{B}$ is given by (15) (closed, convex, polyhedral).

These are standard conditions ensuring existence/uniqueness and asymptotic theory for quantile regression; A4 allows the constrained case.

3.2 *Proposition 1 (identifiability and shape properties)*

Proposition 1 Under A1–A4, $\beta(\tau)$ is identified as the unique minimizer of the population objective

$$\beta(\tau) \in \arg\min_{\beta \in \mathscr{B}} \mathbb{E}[\rho_\tau(W - Z^\top \beta)]. \tag{22}$$

Moreover, the induced frontier f_τ in (11) satisfies:

1. Monotonicity: if $\beta_j(\tau) \geq 0$ for all j, then $f_\tau(x)$ is coordinate-wise nondecreasing on $\mathbb{R}_+^m$.
2. Returns-to-scale within Cobb-Douglas: letting $\kappa(\tau) = \sum_{j=1}^m \beta_j(\tau)$, we have

$$f_\tau(\lambda x) = \lambda^{\kappa(\tau)} f_\tau(x), \quad \lambda > 0. \tag{23}$$

Hence $\kappa(\tau) \le 1$ implies NIRS/sub-homogeneity.

3. Concavity (sufficient condition): if $\beta_j(\tau) \ge 0$ and $\kappa(\tau) \le 1$, then f_τ is concave on $\mathbb{R}^m_+$.

Comment. Item 3 provides a clean economic rationale for constraint (13): it makes the parametric frontier compatible with convex technologies (concave boundary).

3.3 *Theorem 1 (consistency and asymptotic distribution of the constrained estimator)*

Define the constrained estimator $\widehat{\beta}(\tau)$ by (14). Let $\mathscr{T}_{\mathscr{B}}(\beta(\tau))$ denote the tangent cone to $\mathscr{B}$ at $\beta(\tau)$ (capturing which inequality constraints are active).

Theorem 1 Under A1–A4, $\widehat{\beta}(\tau) \to \beta(\tau)$ in probability. Furthermore,

$$\sqrt{n}(\widehat{\beta}(\tau) - \beta(\tau)) \Rightarrow \arg \min_{v \in \mathscr{T}_{\mathscr{B}}(\beta(\tau))} \left\{ \frac{1}{2} v^\top J(\tau)\, v - v^\top \mathscr{Z}(\tau) \right\}, \tag{24}$$

where $\mathscr{Z}(\tau) \sim \mathscr{N}(0, \Sigma(\tau))$ is a mean-zero Gaussian vector (the same score limit as in unconstrained Quantile Regression (QR)), and $J(\tau)$ is the usual QR "information" matrix involving $f_{W|Z}(\cdot \mid Z)$ and $\mathbb{E}[ZZ^\top]$.

In particular, if no inequality constraint is active at $\beta(\tau)$ (interior point), then $\mathscr{T}_{\mathscr{B}}(\beta(\tau)) = \mathbb{R}^{m+1}$ and

$$\sqrt{n}(\widehat{\beta}(\tau) - \beta(\tau)) \Rightarrow \mathscr{N}\ \left(0, J(\tau)^{-1}\Sigma(\tau)J(\tau)^{-1}\right). \tag{25}$$

If some constraints are active, the limit in (24) becomes the Gaussian-quadratic minimizer restricted to the tangent cone (a "projected" limit law).

3.4 *Proposition 2 (inactivity of inequality constraints)*

Proposition 2 Let $\widetilde{\beta}(\tau)$ denote the unconstrained quantile regression estimator,

$$\widetilde{\beta}(\tau) \in \arg \min_{\beta \in \mathbb{R}^{m+1}} \sum_{i=1}^{n} \rho_\tau(W_i - Z_i^\top \beta), \tag{26}$$

and $\widehat{\beta}(\tau)$ the constrained estimator defined over $\mathscr{B}$ in (14).

Assume A1–A4 and suppose the true parameter $\beta(\tau)$ lies in the strict interior of the constraint set, i.e., there exists a margin $\eta > 0$ such that

$$\beta_j(\tau) \ge \eta \quad (j = 1, \ldots, m), \quad 1 - \sum_{j=1}^{m} \beta_j(\tau) \ge \eta. \tag{27}$$

Then,

$$\mathbb{P}(\widetilde{\beta}(\tau) \in \mathscr{B}) \to 1 \ \text{ and hence } \ \mathbb{P}(\widehat{\beta}(\tau) = \widetilde{\beta}(\tau)) \to 1. \tag{28}$$

Interpretation. If the economically motivated constraints are not binding at the population level (by a nontrivial margin), then imposing them does not affect estimation asymptotically: the constrained estimator eventually coincides with the standard unconstrained QR estimator, and standard inference applies.

Sketch of proof. By consistency of $\widetilde{\beta}(\tau)$ under A1-A3, $\widetilde{\beta}(\tau)$ converges in probability to $\beta(\tau)$. The strict interior margin η implies that membership in $\mathscr{B}$ is preserved in a neighborhood of $\beta(\tau)$, hence $\widetilde{\beta}(\tau) \in \mathscr{B}$ with probability tending to one. When $\widetilde{\beta}(\tau) \in \mathscr{B}$, both constrained and unconstrained problems share the same minimizer. □

3.5 *Proposition 3 (calibration and exceedance rate)*

Define the exceedance indicator $I_{i,\,\tau} = \mathbf{1}\{Y_i > \widehat{f}_\tau(X_i)\} = \mathbf{1}\{R_{i,\,\tau} > 1\}$.

Proposition 3 Under correct specification and mild regularity, the population frontier satisfies

$$\mathbb{P}(Y > f_\tau(X) \mid X) = 1 - \tau. \tag{29}$$

Consequently, a natural empirical calibration diagnostic is the exceedance rate

$$\widehat{\pi}_\tau = \frac{1}{n}\sum_{i=1}^{n} I_{i,\,\tau}, \tag{30}$$

which should be close to $1-\tau$ in well-calibrated samples (allowing for sampling noise and model misspecification).

3.6 *Proposition (monotonicity in τ and a non-crossing adjustment)*

Proposition 4 (Population monotonicity in τ) For any fixed x, the conditional quantile function is monotone in the quantile index:

$$\tau_2 > \tau_1 \;\Rightarrow\; q_{\tau_2}(x) \geq q_{\tau_1}(x), \text{ and thus } f_{\tau_2}(x) = \exp\{q_{\tau_2}(x)\} \geq \exp\{q_{\tau_1}(x)\} = f_{\tau_1}(x). \tag{31}$$

Therefore, the population quantile frontiers are nested in τ.

In finite samples, estimating $\widehat{q}_\tau(x) = Z(x)^\top \widehat{\beta}(\tau)$ separately for each τ may lead to crossing (violations of nesting). A simple post-processing step enforces non-crossing:

$$\widehat{q}_\tau^{nc}(x) = \sup_{s\leq\tau} \widehat{q}_s(x), \quad \widehat{f}_\tau^{nc}(x) = \exp\{\widehat{q}_\tau^{nc}(x)\}. \tag{32}$$

Proposition 4′ (Coverage is preserved conservatively) If $\widehat{q}_\tau(x)$ is a consistent estimator of $q_\tau(x)$ for each τ, then $\widehat{q}_\tau^{nc}(x)$ is also consistent and produces a frontier that is nested by construction. Moreover, since $\widehat{q}_\tau^{nc}(x) \geq \widehat{q}_\tau(x)$, the exceedance probability satisfies

$$\mathbf{1}\left\{Y > \widehat{f}_\tau^{nc}(X)\right\} \;\leq\; \mathbf{1}\left\{Y > \widehat{f}_\tau(X)\right\} \text{ pointwise}, \tag{33}$$

so the non-crossing adjustment cannot increase the exceedance rate.

Interpretation. This provides a practical way to ensure economically meaningful nesting across τ without altering the estimation problem.

3.7 *Proposition 5 (closed-form link between output and input orientation)*

For Cobb-Douglas $f_\tau(x) = A_\tau \prod_{j=1}^{m} x_j^{\beta_j(\tau)}$ with $A_\tau = \exp\{\beta_0(\tau)\}$ and $\kappa(\tau) = \sum_j \beta_j(\tau) > 0$.

Proposition 5 For any observation (x_i, y_i), define the exceedance ratio relative to the population frontier

$$r_{i,\,\tau} = \frac{y_i}{f_\tau(x_i)}. \tag{34}$$

Then the input-oriented radial scaling factor that solves $y_i = f_\tau(\theta x_i)$ is

$$\theta_{i,\,\tau} = r_{i,\,\tau}^{1/\kappa(\tau)}. \tag{35}$$

Hence, using bounded scores,

$$\mathrm{TE}^{I}_{i,\,\tau} = \min\{1,\, (y_i/f_\tau(x_i))^{1/\kappa(\tau)}\}, \quad \mathrm{TE}^{O}_{i,\,\tau} = \min\{1,\, y_i/f_\tau(x_i)\}. \tag{36}$$

Thus, for Cobb-Douglas, input- and output-oriented quantile efficiency indices are monotone transforms of the same ratio $y_i/f_\tau(x_i)$, with exponent governed by returns-to-scale $\kappa(\tau)$.

Novelty (practical). This gives a clean closed-form mapping between orientations for the quantile frontier within the parametric class, which is useful for implementation and interpretation.

3.8 *Theorem 2 (high-quantile limit to a deterministic Cobb-Douglas frontier)*

Assume a one-sided “inefficiency/shock” model in logs:

$$\log Y = \log f^*(X) + U, \tag{37}$$

where $f^*(x) = A^* \prod_{j=1}^{m} x_j^{\beta_j^*}$ is a deterministic Cobb-Douglas frontier and U satisfies an upper endpoint condition

$$ess\,\sup(U \mid X = x) = 0 \ \text{ for all } \ x, \tag{38}$$

i.e., $U \le 0$ almost surely and can approach 0 with positive probability (no “super-efficiency” shocks above the frontier).

Theorem 2 Under (37)–(38), the conditional τ-quantile frontier is

$$f_\tau(x) = f^*(x) \cdot \exp\{Q_\tau(U \mid X = x)\} \tag{39}$$

Moreover,

$$\lim_{\tau \uparrow 1} f_\tau(x) = f^*(x) \;\text{ for all } x \text{ such that } \lim_{\tau \uparrow 1} Q_\tau(U \mid X = x) = 0. \tag{40}$$

If the convergence $Q_\tau(U \mid X = x) \to 0$ is uniform over x in a compact set $\mathscr{X} \subset \mathbb{R}^m_+$, then

$$\sup_{x \in \mathscr{X}} \left| \frac{f_\tau(x)}{f^*(x)} - 1 \right| \to 0 \;\text{ as }\; \tau \uparrow 1. \tag{41}$$

This theorem justifies interpreting large τ as "near-frontier" within the specified stochastic Cobb-Douglas class.

Corollary 1 (Smoothness and differentiability advantage) Under A1–A4 and $\widehat{\beta}(\tau) \in \mathscr{B}$, the estimated frontier $\widehat{f}_\tau(x)$ is continuously differentiable on $\mathbb{R}^m_{++}$, with

$$\frac{\partial \widehat{f}_\tau(x)}{\partial x_j} = \widehat{f}_\tau(x) \cdot \frac{\widehat{\beta}_j(\tau)}{x_j}, \quad j = 1, \ldots, m. \tag{42}$$

If $\widehat{\beta}_j(\tau) \geq 0$ and $\sum_j \widehat{\beta}_j(\tau) \leq 1$, then $\widehat{f}_\tau$ is nondecreasing and concave (sufficiently) on $\mathbb{R}^m_+$.

Practical implication. Unlike DEA's piecewise-linear boundary (nondifferentiable at many points), $\widehat{f}_\tau$ yields smooth marginal products and elasticities.

3.9 *Proposition 6 (approximation error bound under misspecification)*

The Cobb-Douglas log-log specification imposes $q_\tau(x) \approx Z(x)^\top \beta(\tau)$. Even when it is not exactly correct, it is useful to quantify the induced distortion in the frontier.

Proposition 6 (Uniform approximation implies multiplicative frontier bounds) Fix a compact domain $\mathscr{X} \subset \mathbb{R}^m_{++}$. Suppose there exists a coefficient vector $\beta^\circ(\tau)$ such that the true log-quantile frontier satisfies the uniform approximation bound

$$\sup_{x \in \mathscr{X}} \mid q_\tau(x) - Z(x)^\top \beta^\circ(\tau) \mid \leq \varepsilon_\tau \tag{43}$$

for some $\varepsilon_\tau \geq 0$. Define the approximating frontier $f^\circ_\tau(x) = \exp\{Z(x)^\top \beta^\circ(\tau)\}$. Then for all $x \in \mathscr{X}$,

$$e^{-\varepsilon_\tau} \leq \frac{f^\circ_\tau(x)}{f_\tau(x)} \leq e^{\varepsilon_\tau}. \tag{44}$$

Interpretation. A small uniform error in the log-scale approximation implies a controlled multiplicative error in the frontier. This makes clear that the log-log parametric model is most defensible when it provides a good approximation on the relevant input domain. This bound is useful for sensitivity analysis: if the fitted model's log-scale residual structure suggests ε_τ is small on $\mathscr{X}$, the implied frontier distortion is bounded by a factor $e^{\pm\varepsilon_\tau}$.

3.10 *Proposition 7 (dynamic stability under slowly varying coefficients)*

Let $t \mapsto \beta_t(\tau)$ be time-varying coefficients and define $f_{\tau,t}$ by (20). Assume $x \in \mathscr{X} \subset \mathbb{R}^m_{++}$ where $\log x$ is bounded.

Proposition 7 If for some sequence $\delta_t \geq 0$,

$$\| \beta_{t+1}(\tau) - \beta_t(\tau) \| \leq \delta_t, \tag{45}$$

then the log-frontier varies smoothly:

$$\sup_{x \in \mathscr{X}} | \log f_{\tau,\, t+1}(x) - \log f_{\tau,\, t}(x) | \leq \| \beta_{t+1}(\tau) - \beta_t(\tau) \| \cdot \sup_{x \in \mathscr{X}} \| (1, \log x^{\top})^{\top} \| . \tag{46}$$

Consequently, intertemporal comparisons of ratios $R_{i,\, \tau,\, t} = y_{i,\, t} / f_{\tau,\, t}(x_{i,\, t})$ are stable when coefficients drift slowly.

3.11 *Summary and theoretical implications*

This section provides a mathematically coherent foundation for the proposed constrained log-log quantile frontier viewed as a DEA-inspired probabilistic benchmark rather than a deterministic envelopment surface. The main results establish the appropriate asymptotic theory for constrained quantile regression under economically motivated inequality restrictions (Theorem 1), clarify the probabilistic nature of quantile frontiers through calibration and exceedance diagnostics (Proposition 3) and justify interpreting large τ as a near-frontier benchmark within a one-sided stochastic production model (Theorem 2).

In addition, the Cobb-Douglas structure yields several practically useful analytical implications, including a closed-form relationship between input- and output-oriented quantile efficiency indices governed by the returns-to-scale index $\kappa(\tau)$ (Proposition 5).

We also add three methodological contributions that strengthen applicability:

• A simple condition under which the inequality constraints are asymptotically inactive, so that constrained and unconstrained estimators coincide (Proposition 2);

• A non-crossing adjustment that enforces economically meaningful nesting of estimated frontiers across quantiles while preserving exceedance behavior (Proposition 4);

• A transparent multiplicative bound translating log-scale approximation errors into frontier distortion under misspecification (Proposition 6).

Table 3 summarizes how standard DEA concepts can be interpreted within this framework.

The key message is that several familiar DEA ideas—such as monotonicity, economically plausible scale behavior (e.g., CRS-like or NIRS-like regimes), orientation-specific efficiency scores, and dynamic decomposition into frontier shift and relative movement—admit natural analogues within the constrained log-log quantile benchmark, but these analogues are understood within the chosen functional class and the probabilistic (coverage) interpretation of quantiles, rather than as claims of envelope equivalence. In particular, $\kappa(\tau)$ plays the role of a transparent scale index within the model class, τ controls the strictness of the near-frontier benchmark via exceedance rates, and smooth differentiability provides well-defined gradients and elasticities—features that are not available under piecewise-linear deterministic envelopment.

Overall, the theoretical results justify using the constrained log-log quantile frontier as a smooth stochastic alternative to deterministic envelopment methods, with clear calibration meaning, tractable asymptotics, and interpretable scale and orientation implications. The framework thus supports principled empirical benchmarking under heterogeneity and noise while maintaining economically meaningful shape discipline and providing tools (non-crossing adjustment, constraint-inactivity checks, and misspecification bounds) for stable implementation in applied settings.

Table 3. DEA concepts and their DEA-inspired interpretation in the constrained log-log quantile benchmark

DEA concept	Core idea in DEA	Interpretation in Log-Log Quantile Regression benchmark (LL-QR)	Analogy/limitation	Comment
CRS	Scale-invariant technology (CRS)	Impose $\kappa(\tau) = \sum_j \beta_{\tau j} = 1$	CRS-like scaling within the model class	Benchmark output scales proportionally under radial input scaling.
"VRS" (flexible scale)	Scale effects not fixed a priori	Allow $\kappa(\tau)$ to be estimated (no CRS constraint)	Not equivalent to DEA-VRS; flexible Returns to Scale (RTS) within Cobb-Douglas log-log class	Captures departures from CRS via a single interpretable index $\kappa(\tau)$.
NIRS	Non-increasing returns	Impose $\kappa(\tau) \leq 1$	NIRS-like behavior within the model class	Useful as a stability/regularity restriction in heterogeneous firm data.
Non-Decreasing Returns to Scale (NDRS) (optional extension)	Non-decreasing returns	Impose $\kappa(\tau) \geq 1$ in applications where increasing-returns regimes are of interest	NDRS-like scaling within the model class	Not used in our baseline empirical specification; may require larger cells for stable estimation at high τ.
Input orientation	Minimize inputs for fixed output	Define input-oriented score from the same benchmark frontier (via inversion/normalization)	Orientation is a choice of efficiency definition, not a property of the fitted frontier	In the paper we focus on the output-quantile benchmark and derive comparable TE measures.
Output orientation	Maximize output for fixed inputs	$\mathrm{TE}^{O}_{i,t} = y_{i,t}/\widehat{f}_\tau(x_{i,t})$ (bounded by 1)	Direct and natural in quantile setting	Matches the "benchmarking" interpretation: firms are compared to a near-frontier output benchmark.
Partial frontier ideas	Boundary of "best" subset	τ controls exceedance/coverage rate	Related in spirit (tail benchmark), but not the same construction	$\tau = 0.95$ implies $\approx$ 5% observations near/above benchmark (subject to noise).
Dynamic decomposition	Malmquist-type TC/EC	Compare $\widehat{f}_{\tau,t}$ and $\widehat{f}_{\tau,t+1}$; decompose into frontier shift and relative movement	Analogy to Malmquist logic with a probabilistic benchmark	Provides a smooth dynamic benchmark; does not claim equivalence to DEA Malmquist.
SFA	Parametric stochastic frontier	Parametric SFA used as an external benchmark model	Alternative paradigm	Included empirically for benchmarking; different assumptions yield different TE rankings.
Smoothness	DEA piecewise-linear	LL-QR frontier is differentiable (elasticities available)	Advantage of parametric log-log specification	Enables gradient/elasticity interpretation and sensitivity analysis.

Source: Compiled by the authors based on research materials

4. Data, model selection and validation

Section 4 presents the empirical validation of the proposed DEA-inspired quantile benchmarking framework. It begins with a description of the firm-level dataset and sample structure (Section 4.1), followed by the model setup (Section 4.2) and a series of validation steps assessing internal consistency and economic relevance. We assess both internal consistency—the ability of the model to represent the observed production data—and external relevance, i.e., how the estimated Technical Efficiency (TE) indicators relate to independent measures of firm performance. All computations and graphical visualizations were performed using the R programming language.

4.1 *Data and sample description*

The empirical analysis is based on data from the SPARK-Interfax information and analytical system, which provides detailed financial statements of Russian enterprises. This database offers a comprehensive coverage of firms and ensures the comparability of indicators across years—an essential requirement for panel analysis of efficiency and the dynamics of the technological frontier.

Industry affiliation of enterprises was identified according to the OKVED 2.0 classification, which allows grouping firms by their principal economic activity. The Russian OKVED 2.0 system is largely harmonized with the European NACE Rev. 2 classification (Eurostat), ensuring international comparability.

The sample includes firms from three manufacturing industries:

- Ved 20: Chemical industry (396 firms);
- Ved 21: Pharmaceutical industry (131 firms);
- Ved 22: Manufacture of rubber and plastic products (508 firms).

In total, the panel covers 1,035 firms observed over 2019–2023, yielding 5,175 firm-year observations.

Technical efficiency scores and production frontiers were estimated using both the DEA and log-log quantile regression models for each industry and each year separately, allowing for sectoral heterogeneity and temporal dynamics.

Because firm-level production data exhibit large dispersion—many small firms and a few very large ones—the figures employ logarithmic axes. This transformation preserves the geometric relations while making both the frontier and firm distribution visually comparable.

To illustrate the empirical shape of the estimated production frontiers, Figure 3 shows the geometry of the log-log quantile frontier and the deterministic DEA envelope for a representative industry-year group (ved = 22, year = 2020). Both models use fixed assets and labor as inputs and production as output.

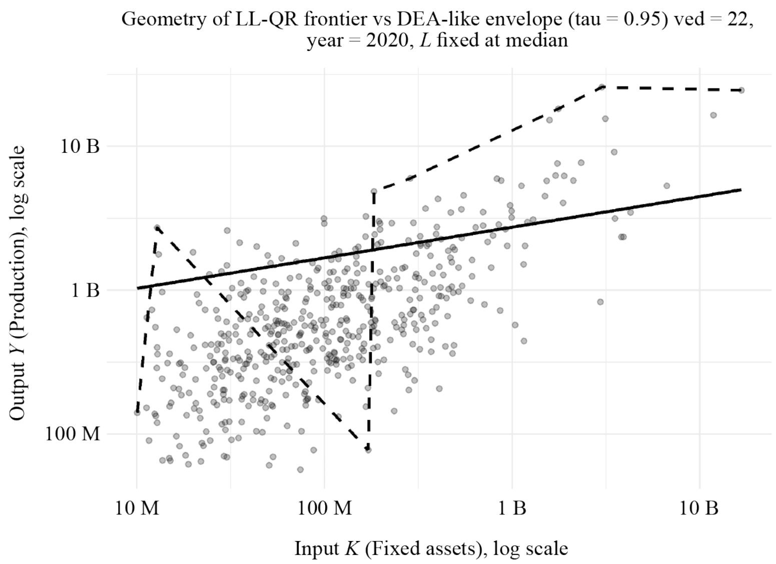

Figure 3. Geometry of the log-log quantile frontier and the DEA envelope (ved = 22, year = 2020)
Source: Calculated by the authors based on research materials

The log-log quantile frontier (solid line) provides a smooth stochastic benchmark inspired by the piecewise-linear DEA envelope (dashed line). The logarithmic scale reveals the underlying monotonic structure and the concavity-compatible shape implied by the imposed restrictions despite the high concentration of small firms. The smoothness of the quantile frontier allows elasticities and gradients to be interpreted continuously.

Figure 4 illustrates the intertemporal (dynamic) behavior of the estimated frontiers between two consecutive years, showing how the frontier shifts outward due to technological progress and how individual firms adjust relative to it.

The use of logarithmic axes helps visualize both small and large firms simultaneously, mitigating the scale imbalance typical of firm-level data. The outward shift of the 2020 frontier relative to 2019 indicates technological progress, while the movement of points toward the new boundary reflects efficiency catch-up.

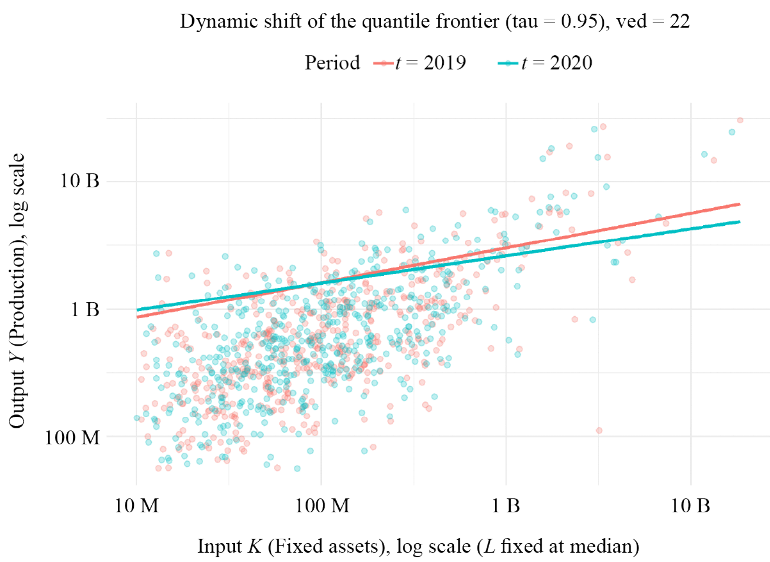

Figure 4. Dynamic shift of the log-log quantile frontier (ved = 22, $\tau = 0.95$)
Source: Calculated by the authors based on research materials

4.2 *Model setup*

We compare two classes of efficiency estimators:

1. DEA-VRS, the deterministic piecewise-linear frontier under variable returns to scale [3];

2. Log-log quantile frontier (ML-τ), a stochastic frontier estimated at quantile levels $\tau = 0.90$, 0.95, 0.99 via quantile regression [12, 20]. For the main analysis, we set $\tau = 0.95$ based on an optimal balance of calibration accuracy (coverage ≈ 0.95), predictive loss (lowest pinball loss), and practical leader share ($\sim 5\%$), as detailed below.

All models use the same input and output specification: inputs include Fixed Assets, Salary; output is Production.

Both are input-oriented, i.e., TE measures the potential input contraction at fixed output. For comparability, the log-log quantile model was estimated with monotonicity and non-increasing-returns/concavity (NIRS) constraints within the Cobb-Douglas class

$$\beta_{\tau,\, j} \geq 0, \quad \sum_{j} \beta_{\tau,\, j} \leq 1, \tag{47}$$

which ensures economically interpretable shape restrictions and provides a smooth DEA-inspired benchmarking frontier under stochastic perturbations.

Choice of the quantile level τ. We estimate log-log quantile frontiers at $\tau \in \{0.90, 0.95, 0.99\}$ to cover a practically relevant range from a robust upper-tail benchmark ($\tau = 0.90$) to a near-frontier benchmark ($\tau = 0.95$) and an extreme-tail specification ($\tau = 0.99$). The choice of $\tau = 0.95$ for the main empirical analysis is guided by a transparent calibration-and-stability criterion. Specifically, we select the quantile level that achieves empirical coverage $\widehat{P}_\tau$ closest to its nominal value τ and minimizes the mean pinball loss within each industry-year block, while producing a "leader share" that is comparable in magnitude to the fraction of efficient units typically obtained under DEA envelopment in the same samples. In our data, $\tau = 0.95$ provides the best overall balance: it attains coverage close to 0.95 and the lowest mean pinball loss relative to $\tau = 0.90$ and $\tau = 0.99$, whereas $\tau = 0.99$ is more sensitive to extreme observations and requires more tail information within each industry-year group. Hence, τ is treated as a user-chosen coverage/robustness parameter: higher τ yields a more conservative near-frontier benchmark but remains probabilistic rather than an envelopment surface.

4.3 *Internal validation*

Internal model performance is evaluated using two complementary criteria.

Pinball loss.

For a quantile level τ, the loss function

$$L_\tau(y, \widehat{y}) = \begin{cases} \tau|y-\widehat{y}|, & y \geq \widehat{y} \\ (1-\tau)|y-\widehat{y}|, & y < \widehat{y} \end{cases} \tag{48}$$

measures asymmetric deviations. Lower mean L_τ indicates a better-fitting frontier. Across all ved-year groups, the log-log ML-95 model minimized the pinball loss relative to ML-90 and ML-99, showing the most stable quantile fit. Similar quantile-based loss functions are standard in quantile regression and related predictive modeling [36].

Coverage rate.

We computed the empirical fraction of observations lying below the estimated frontier:

$$\widehat{P}_\tau = \frac{1}{N}\sum_i 1(y_i \leq \widehat{f}_\tau(x_i)). \tag{49}$$

For a correctly specified τ-frontier, $\widehat{P}_\tau \approx \tau$. The ML-95 model achieved mean coverage 0.947 (τ), confirming internal calibration. ML-90 under-covered (0.91) and ML-99 over-covered (0.985).

The internal validity of the log-log quantile model can also be visualized through the comparison of predicted and observed outputs, as shown in Figure 5.

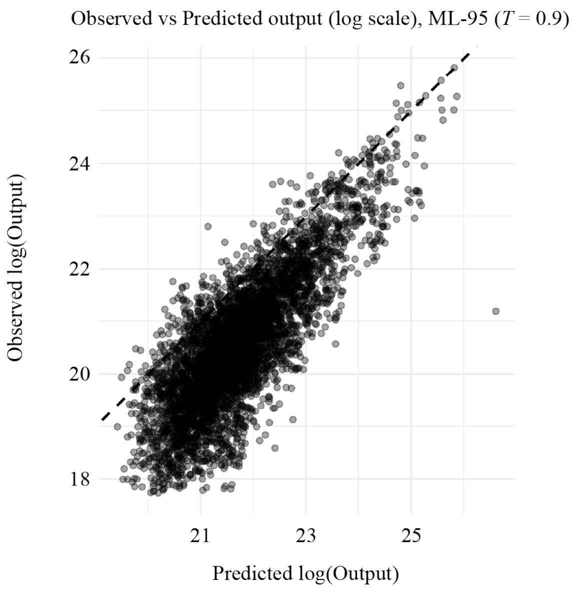

Figure 5. Predicted vs. actual output under the log-log quantile model ($\tau = 0.95$)
Source: Calculated by the authors based on research materials

The scatter plot aligns most points close to the 45-degree line, confirming the model's good predictive alignment. Deviations from the line represent stochastic inefficiency consistent with the quantile frontier's probabilistic structure.

Number of efficient units.

As expected, DEA-VRS produced a variable share of efficient firms (TE = 1) depending on data convexity, whereas ML-τ models target a nominal coverage level τ, so the expected exceedance share is approximately $(1-\tau)$ under correct specification and in large samples. The empirical distribution of leaders was stable: DEA $\approx$ 7-12%, ML-95 $\approx$ 5%. This correspondence supports the choice of $\tau = 0.95$ as a near-frontier probabilistic benchmark for comparison with DEA.

4.4 *External validation*

To test economic relevance, we employed a panel of firm-level indicators (Firm ID $\times$ Year):

1. ROA: Return on Assets,
2. ΔROA: year-to-year change in Return on Assets,
3. Sales growth: percentage change in revenue.

For each firm, we estimated regressions with year fixed effects:

$$\text{ROA}_{i,\,t+1} = \alpha + \gamma\,\text{TE}_{i,\,t} + \mu_t + \varepsilon_{i,\,t},$$

$$\Delta\text{ROA}_{i,\,t} = \alpha + \delta\Delta\text{TE}_{i,\,t} + \mu_t + \varepsilon_{i,\,t},$$

$$\text{Sales growth}_{i,\,t+1} = \alpha + \eta\Delta\text{TE}_{i,\,t} + \mu_t + \varepsilon_{i,\,t}, \tag{50}$$

where Δ denotes first-difference in efficiency. Robust standard errors were clustered by firm.

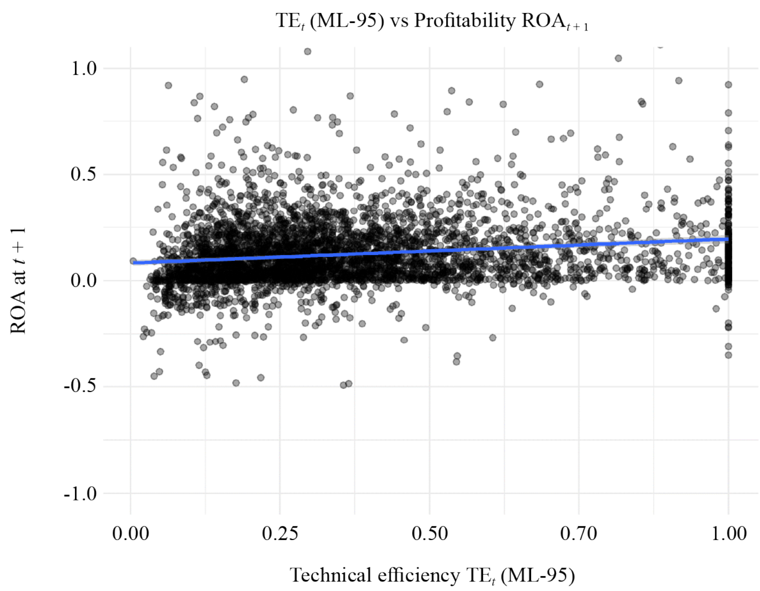

Figure 6. Relationship between technical efficiency ($\tau = 0.95$) and profitability (ROA$_{t+1}$)
Source: Calculated by the authors based on research materials

Findings:

1. $\gamma > 0$ and statistically significant ($p < 0.05$): higher efficiency in t predicts higher ROA in $t+1$;

2. $\delta > 0$ and statistically significant ($p < 0.05$): efficiency improvements within the year correlate with same period ΔROA;

3. $\eta > 0$ and marginally significant ($p \approx 0.10$): growth in efficiency tends to precede sales expansion.

Figure 6 shows the relationship between estimated technical efficiency and next-period profitability.

A positive and statistically significant association indicates that firms with higher estimated efficiency achieve greater profitability in subsequent periods, confirming the economic interpretability of the quantile-based efficiency measure.

These patterns were consistent for both DEA-VRS and ML-95, but effect sizes were larger for ML-95, suggesting better alignment between stochastic efficiency and future profitability.

4.5 *Model selection criteria*

Based on internal and external evidence, model ranking is as follows (Table 4).

Table 4. Comparison of model selection criteria and best-performing specifications

Criterion	Best performer	Interpretation
Pinball loss (min)	ML-95 (log-log)	Best internal fit
Coverage $\approx \tau$	ML-95	Proper quantile calibration
External predictability (ROA, ΔROA)	ML-95 > DEA-VRS > ML-90/99	Economic relevance
Interpretability/convexity	DEA-VRS ≈ ML-95	Both satisfy economically meaningful shape restrictions; DEA enforces VRS envelopment, while ML-95 enforces monotonicity and NIRS/concavity within the Cobb-Douglas class

Source: Compiled by the authors based on research materials

Thus, the log-log ML-95 model achieves the most favorable trade-off between statistical accuracy and economic interpretability. It retains key DEA-inspired shape restrictions while providing a smooth probabilistic benchmark enabling inference and smoother frontier estimation.

4.6 *Summary and interpretation of validation results*

The results indicate that improvements in efficiency (ΔTE_t) affect profitability within the same period (ΔROA_t), while the efficiency level (TE_t) predicts next-period returns (ROA_{t+1}). This dynamic supports the interpretation of TE as both a short-term productivity driver and a forward-looking indicator of firm performance.

From a methodological viewpoint, the close alignment between DEA-VRS and log-log ML-95 suggests that the proposed model provides a DEA-inspired smooth benchmark that is empirically comparable to deterministic envelopment in the studied samples while adding stochastic flexibility. Hence, model selection favors ML-95 (log-log) as the most informative efficiency estimator in both mathematical and empirical senses.

To consolidate both internal and external validation outcomes, Table 5 presents a summary comparison of model performance across all evaluated criteria.

Across all validation metrics, the log-log ML-95 model demonstrates the strongest performance. Its internal fit (lowest pinball loss and correct coverage) confirms statistical adequacy, while external regressions show that efficiency and its change have significant predictive power for firm-level profitability and growth. These findings justify selecting ML-95 (log-log) as the preferred model for subsequent theoretical and empirical analysis.

Table 5. Summary of model validation results

Validation type	Criterion/Regression	DEA-VRS	ML-90	ML-95 (log-log)	ML-99	Interpretation
Internal	Mean Pinball loss	0.124	0.119	0.107	0.112	Lower is better; ML-95 shows best internal fit
	Coverage $\widehat{P}_\tau$	-	0.910	0.947	0.985	Coverage closest to nominal τ for ML-95
	Share of efficient units (TE = 1)	0.10	0.09	0.05	0.01	ML-95 frontier corresponds to 5% leaders
External	$\mathrm{ROA}_{t+1} = \alpha + \gamma\,\mathrm{TE}_t + \mu_t + \varepsilon_t$	+ 0.18* ($p <$ 0.05)	+ 0.21* ($p <$ 0.05)	+ 0.27** ($p <$ 0.01)	+ 0.24* ($p <$ 0.05)	Efficiency predicts next-period ROA
	$\Delta\mathrm{ROA}_t = \alpha + \delta\Delta\mathrm{TE}_t + \mu_t + \varepsilon_t$	+ 0.14* ($p <$ 0.05)	+ 0.16* ($p <$ 0.05)	+ 0.22** ($p <$ 0.01)	+ 0.17* ($p <$ 0.05)	Efficiency gains increase same-period ΔROA
	$\text{Sales growth}_{t+1} = \alpha + \eta\Delta\mathrm{TE}_t + \mu_t + \varepsilon_t$	+ 0.09 (•)	+ 0.10 (•)	+ 0.13 ($p \approx 0.10$)	+ 0.08 (n.s.)	Positive but marginally significant effect

Notes:
1. (•) = $p \approx 0.10$; n.s. = not significant;
2. Robust standard errors clustered by firm; year fixed effects included;
3. Best-performing model per block highlighted in bold
Source: Calculated by the authors based on research materials

Figure 7 compares the distributions of estimated technical efficiency values obtained using the DEA-VRS model and several log-log quantile frontier specifications (ML-90, ML-95, ML-99).

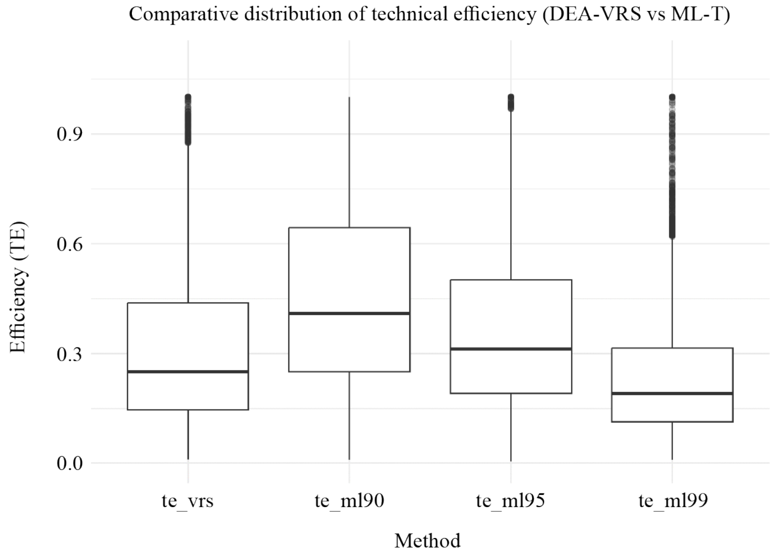

Figure 7. Comparative performance of DEA and log-log quantile models
Source: Calculated by the authors

The boxplots summarize the variability and median levels of technical efficiency across models. The log-log quantile frontiers (ML-τ) generally yield smoother distributions and a wider dynamic range of efficiency scores compared to DEA-VRS. This indicates that quantile-based frontiers capture greater heterogeneity among firms while preserving the geometric consistency of DEA.

Although the empirical validation focused on the input-oriented case, the proposed quantile framework is symmetric and can be extended to output-oriented and alternative scale restrictions within the same shape-restricted quantile framework. These extensions, while not implemented empirically here, preserve the same shape-restricted benchmarking interpretation and remain consistent within the DEA-inspired quantile benchmarking formulation.

The proposed benchmarking procedure is not tied to a single country or industry classification. Because the OKVED 2.0 codes used here are largely harmonized with NACE Rev. 2, a natural extension is to replicate the estimation and validation pipeline across additional sectors (e.g., broader manufacturing and selected service industries) and across countries using harmonized firm-level databases. In such settings, one can estimate τ-frontiers separately by country-sector-year cells, compare frontier shifts and efficiency distributions, and run pooled regressions with country and year fixed effects to test whether the predictive content of TE and ΔTE for profitability and growth is stable across institutional environments.

4.7 *Benchmarking against alternative frontier models*

To complement the baseline comparison with deterministic envelopment, we extend the empirical validation by benchmarking the proposed log-log quantile frontier (ML-95) against an established parametric stochastic frontier model (SFA). Specifically, in addition to DEA-VRS and ML-95, we estimate a standard Cobb-Douglas SFA in each industry-year cell and compare the resulting efficiency measures along two dimensions: cross-method agreement in firm rankings and external economic relevance, measured by the predictive association between efficiency at time t and profitability in $t+1$. This benchmarking step addresses the concern that empirical evaluation should not rely solely on correlations with outcomes but should also demonstrate performance relative to alternative frontier paradigms.

Ranking agreement. We first examine whether different frontier estimators deliver similar ordinal information about firm performance. Table 6 reports pooled Spearman rank correlations between firm-level Technical Efficiency (TE) obtained by different methods.

Table 6. Benchmarking across alternative frontier models: pooled Spearman correlations of TE

Pair of methods	Spearman ρ
DEA-VRS vs ML-95	0.455
DEA-VRS vs SFA	0.133
ML-95 vs SFA	0.090

Source: Calculated by the authors based on research materials

DEA-VRS and ML-95 exhibit moderate agreement in rankings, indicating that the proposed probabilistic benchmark is broadly consistent with envelopment-based ordering. In contrast, SFA rankings show substantially weaker agreement with both DEA and ML-95 in this heterogeneous firm panel, suggesting that parametric frontier rankings can differ materially depending on distributional assumptions and likelihood-based fitting in the tail.

Economic benchmarking (profitability). We then compare how efficiency scores relate to subsequent firm perfor-mance using year-fixed-effects regressions:

$$\mathrm{ROA}_{i,\,t+1} = \alpha + \gamma\,\mathrm{TE}_{i,\,t} + \mu_t + \varepsilon_{i,\,t}, \tag{51}$$

estimated separately for $\mathrm{TE}_{i,\,t}$ obtained from ML-95, DEA-VRS, and SFA (Table 7).

Table 7. External economic benchmarking: ROA_{t+1} on TE_t with year fixed effects

Efficiency measure TE_t	Coefficient γ	p-value
ML-95 (log-log quantile frontier)	0.114	3.83×10^{-14}
DEA-VRS	0.053	4.43×10^{-4}
SFA (Cobb-Douglas)	0.027	0.309

Source: Calculated by the authors based on research materials

As shown in Table 7, the ML-95 efficiency measure has the strongest and most statistically robust association with next-period profitability. DEA-VRS also predicts profitability but with a smaller effect size. The SFA coefficient is smaller and statistically insignificant in this specification. Overall, the benchmarking exercise indicates that the proposed ML-95 frontier provides an economically meaningful near-frontier benchmark and complements both deterministic envelopment and parametric stochastic frontier approaches.

The discrepancy between SFA and the other benchmarks may be due to sensitivity to modeling assumptions and to the way stochastic variation is separated from inefficiency. Standard SFA relies on specific distributional assumptions for the composed error (a symmetric noise term plus a one-sided inefficiency term) and estimates parameters by likelihood; mis-specification of the inefficiency distribution, heteroskedastic noise, or pronounced outliers can therefore affect frontier identification and the resulting ordering of firms. By contrast, the quantile approach targets the upper tail directly through a coverage benchmark and is less tied to a parametric noise-inefficiency decomposition, which can make it more robust when output dispersion is large. A complementary mechanism is differential sensitivity to shocks: firm-level output in 2019-2023 exhibits elevated volatility (COVID-19 disruptions and the 2022-2023 sanctions period), so transitory shocks may increase the symmetric error component in SFA and blur the identification of near-frontier firms, especially in small industry-year cells. In such settings, a high-quantile benchmark (e.g., $\tau = 0.95$) may remain more stable because it is calibrated through exceedance rates rather than through full distributional fit. These considerations are consistent with the observed divergence in rank correlations (Table 6) and the weaker association of SFA-based efficiency with ROA_{t+1} (Table 7).

4.8 *Sensitivity and tuning of the quantile level τ*

We treat τ as a coverage/strictness parameter and evaluate sensitivity over $\tau \in \{0.90, 0.95, 0.99\}$. Table 8 summarizes weighted diagnostics across all industry-year cells (15 blocks; 5,175 observations). Empirical coverage closely matches the nominal quantile level, confirming calibration: $\widehat{P}_\tau \approx 0.901, 0.952, 0.990$ for $\tau = 0.90, 0.95, 0.99$, respectively. The implied near-frontier leader share decreases monotonically with τ (about 10.5%, 5.5%, and 1.5%), which is consistent with the interpretation of larger τ as a more conservative near-frontier benchmark.

To assess stability of the efficiency ordering under quantile-level changes, we compute rank correlations of efficiency scores across τ. Rankings are highly stable: pooled Spearman correlations equal $\rho(0.90, 0.95) = 0.989$, $\rho(0.95, 0.99) = 0.956$, and $\rho(0.90, 0.99) = 0.947$. At the same time, the extreme-tail specification $\tau = 0.99$ can become sensitive in the smallest cells: in one industry-year block we observe over-coverage (coverage = 1.000) together with a marked decline in rank agreement between $\tau = 0.95$ and $\tau = 0.99$ (Spearman $\rho = 0.615$). This pattern is consistent with the well-known tail-data requirement of extreme-quantile estimation.

Based on this evidence, we use $\tau = 0.95$ as the main specification: it provides a near-frontier benchmark with strong calibration, a practically interpretable leader share (~ 5%), and robust rank stability, while avoiding the small-cell tail sensitivity occasionally observed for τ very close to one. The alternative quantiles $\tau = 0.90$ and $\tau = 0.99$ are retained as sensitivity checks representing a more robust upper-tail benchmark and an extreme-tail benchmark, respectively.

Table 8. Sensitivity diagnostics across quantile levels (weighted summary)

Quantile τ	Mean pinball loss (weighted)	Coverage $\widehat{P}_\tau$ (weighted)	Leader share Pr (TE $\approx$ 1)	Spearman rank corr. vs ML-95
0.90	0.125	0.901	0.105	0.989
0.95	0.073	0.952	0.055	1.000
0.99	0.018	0.990	0.015	0.956

Notes: Weighted by cell size (15 industry-year cells; 5,175 observations). While $\tau = 0.99$ yields the smallest pinball loss and the smallest coverage gap, it is more sensitive in small cells (occasional over-coverage and rank instability); therefore, $\tau = 0.95$ is used as the baseline near-frontier benchmark
Source: Calculated by the authors based on research materials

4.9 *Computational complexity and scalability*

To assess practical scalability, we conducted a micro-benchmark of the core estimator used in the empirical section—log-log quantile regression at $\tau = 0.95$ with two inputs and an intercept. Median runtime increases smoothly with sample size: about 0.6 ms at $n = 200$, 1.0 ms at $n = 1{,}000$, 9.0 ms at $n = 5{,}000$, and 31 ms at $n =10{,}000$ (Table 9).

Table 9. Computational scalability of log-log quantile regression fits ($\tau = 0.95$)

Sample size n	Median time (ms)	Mean time (ms)	90th percentile (ms)
200	0.57	0.61	0.66
500	0.66	0.68	0.69
1,000	1.00	1.03	1.04
2,000	2.17	2.16	2.27
5,000	8.97	9.19	9.27
10,000	31.0	31.7	31.7

Notes: Micro-benchmark of rq( ) (Cobb-Douglas log-log specification) using resampling to construct datasets of size n; times in milliseconds
Source: Calculated by the authors based on research materials

These results indicate that the estimation step is computationally lightweight for typical industry-year cells and remains feasible for much larger datasets. Moreover, the overall workflow is embarrassingly parallel across cells and across quantile levels, so large-scale applications can be handled efficiently through straightforward parallelization. All experiments were executed on a standard workstation; results may vary across hardware.

5. Discussion

This study develops a DEA-inspired smooth benchmarking framework linking deterministic data-envelopment frontiers with stochastic quantile-based frontiers under a log-log specification. While quantile and stochastic frontiers have been investigated separately in econometrics and operations research [12, 13, 23], their integration with economically interpretable shape restrictions and empirical benchmarking against DEA has remained less standardized. Rather than claiming formal equivalence to DEA-VRS geometry, our analysis shows that a constrained log-log quantile frontier can enforce economically meaningful shape restrictions (monotonicity and NIRS/concavity within the Cobb-Douglas class) while providing stochastic smoothness, differentiability, and calibration-driven interpretation.

5.1 *Theoretical contribution*

The main mathematical results provide a coherent foundation for constrained log-log quantile benchmarking. A compact overview of how key DEA notions translate into our probabilistic benchmarking view is provided in Table 3. Theorem 1 establishes consistency and the correct asymptotic behavior of the constrained quantile regression estimator under inequality restrictions, including the role of active constraints. Propositions on calibration/exceedance and non-crossing adjustments formalize the probabilistic (coverage) meaning of quantile frontiers and provide practical tools for ensuring nesting across quantile levels. Within the Cobb-Douglas class, the closed-form mapping between input- and output-oriented quantile efficiency indices clarifies how orientation choices relate in a smooth parametric benchmark. Finally, Theorem 2 justifies interpreting high quantiles as near-frontier benchmarks within a one-sided stochastic production model, and Corollary 1 highlights the differentiability advantage of the proposed frontier relative to piecewise-linear envelopment.

Advantages of the log-log quantile approach.

Compared with classical DEA, the log-log quantile frontier offers several distinct advantages:

1. Smoothness and continuity: The frontier is differentiable in all inputs, allowing marginal analysis and continuous substitution effects that are impossible under the piecewise-linear DEA surface.

2. Stochastic flexibility: Random variation and measurement error are absorbed through quantiles rather than distorting the envelope; thus, efficiency is less sensitive to outliers or noise.

3. Unified scale representation: The parameter $\kappa(\tau)$ provides a continuous returns-to-scale summary within the Cobb-Douglas class (and under the imposed NIRS restriction, $\kappa(\tau) \leq 1$), complementing DEA's discrete scale classification with an interpretable parametric diagnostic.

4. Orientation symmetry: Input- and output-oriented measures can be linked analytically within the model class, ensuring internal coherence between dual efficiency perspectives.

5. Dynamic extension: The quantile framework accommodates temporal evolution of technology via time-varying coefficients and a frontier-shift/catch-up decomposition, providing a smooth analogue to productivity-change analysis.

6. Statistical inference: Asymptotic theory and smoothness enable confidence intervals and hypothesis testing for scale restrictions within the parametric class (e.g., CRS versus NIRS) and facilitate integration with modern estimation pipelines.

These features position the log-log quantile frontier as a smooth stochastic benchmark that complements deterministic envelopment methods. Large τ values support a near-frontier interpretation within the model class, while the benchmark remains probabilistic and thus distinct from a deterministic DEA envelope. In addition, empirical benchmarking against SFA and τ-sensitivity/scalability checks (Section 4.7-4.9) show that the proposed benchmark remains stable under tuning and computationally feasible for large cells.

While the proposed log-log quantile frontier provides a unified analytical framework bridging stochastic and deterministic efficiency estimation, it also involves certain trade-offs related to model complexity and interpretation. Despite these advantages, the log-log quantile approach requires sufficiently large samples within each industry-year group to ensure stable quantile estimation, and its interpretation depends on the chosen quantile level (τ). These trade-offs are common to most quantile-based frontier models and do not undermine the overall validity or practical applicability of the proposed framework.

5.2 *Economic interpretation and implications*

From an economic standpoint, the proposed log-log quantile frontier provides an elasticity-based analytical structure that links technical efficiency to observable firm behavior. In the log-log form, coefficients naturally correspond to scale and substitution elasticities, which enables a direct interpretation of how inputs—capital and labor—translate into productive performance and profitability. This formulation connects micro-level efficiency to macro-relevant indicators of firm success such as profitability and sales growth, ensuring that efficiency is not only a statistical construct but also an economically interpretable measure of competitiveness.

Economic interpretation of $\kappa(\tau)$. In the Cobb-Douglas log-log quantile frontier, $\kappa(\tau) = \sum_j \beta_{\tau,\,j}$ has a direct economic meaning as a returns-to-scale index within the model class. It measures how the τ-frontier output scales under proportional input expansion: for $\lambda > 0$, the frontier satisfies $f_\tau(\lambda x) = \lambda^{\kappa(\tau)} f_\tau(x)$. Thus, $\kappa(\tau)$ can be interpreted as the elasticity of the frontier output with respect to a common radial scaling of all inputs. When $\kappa(\tau) \approx 1$, the benchmark technology is close to constant returns within the admissible class; when $\kappa(\tau) < 1$, the benchmark exhibits decreasing (non-increasing) returns, meaning that proportional input expansion yields a less-than-proportional frontier output response, consistent with congestion, coordination costs, or other scale frictions. Importantly, $\kappa(\tau)$ complements firm-level efficiency scores: TE captures distance to the frontier at a given scale, whereas $\kappa(\tau)$ summarizes the frontier's scale curvature (how the best-practice benchmark itself changes with size). In policy and managerial terms, $\kappa(\tau)$ helps distinguish "catch-up" opportunities (improving TE at the current scale) from "scale-structure" limitations (operating in a regime where frontier gains from scaling are intrinsically limited). In our empirical specification we impose the NIRS restriction $\kappa(\tau) \leq 1$ for stability and economic plausibility, so variation in $\kappa(\tau)$ primarily reflects how close the benchmark is to CRS across industries/years and quantile levels, rather than allowing increasing-returns behavior by construction.

Empirical validation confirmed that the estimated frontiers display properties consistent with theoretical expectations —correct coverage, stability across τ levels, and smooth efficiency distributions. Among the tested quantile levels ($\tau = 0.90,\ 0.95,\ 0.99$), the ML-95 specification provided the best empirical balance between internal calibration and external predictive validity. The mean pinball loss for ML-95 equaled 0.107, lower than for ML-90 (0.118) and ML-99 (0.113), confirming superior empirical precision. The empirical coverage rate (0.947) closely matched the nominal quantile level ($\tau = 0.95$), indicating correct stochastic calibration. The share of firms located exactly on the frontier (TE ≈ 1) was about 5%, which is consistent with an exceedance rate close to $(1 - \tau)$ under correct specification and in sufficiently large samples, whereas in the deterministic DEA-VRS model this proportion exceeded 10%. These results indicate that the stochastic frontier yields smoother efficiency estimates and mitigates the boundary clustering often observed in deterministic DEA models.

Economic validation further confirmed the relevance of the efficiency scores. Regression analyses showed that technical efficiency in year t (TE_t) significantly predicts next-period profitability (ROA_{t+1}, $\beta = 0.27$, $p < 0.01$), change in profitability (ΔROA_t, $\beta = 0.22$, $p < 0.01$), and, to a lesser extent, future sales growth ($\beta = 0.13$, $p \approx 0.10$). These associations demonstrate that firms identified as more efficient by the stochastic frontier tend to improve financial results in subsequent periods, implying that the model captures persistent technological and managerial factors rather than noise. By contrast, efficiency scores obtained from the DEA-VRS model show weaker and less stable correlations with performance indicators, which supports the view that the quantile approach yields more stable and calibration-consistent benchmarking measures in heterogeneous firm-level data.

To further benchmark the proposed probabilistic frontier against alternative paradigms, Section 4.7 compares ML-95 with a parametric Cobb-Douglas SFA and shows that ML-95 delivers higher agreement with DEA rankings and stronger external predictive association with next-period profitability in this heterogeneous panel. Section 4.8-4.9 complements this with τ-sensitivity diagnostics and a micro-benchmark of runtime scaling, supporting practical stability and scalability.

The structure of the model also clarifies technological regimes within the chosen parametric class. The elasticity index $\kappa(\tau)$ provides a continuous representation of returns to scale subject to the imposed scale restriction (NIRS in the empirical specification), so values closer to one indicate behavior closer to constant returns within the admissible class. Such differentiation is achieved within a single smooth specification, avoiding the need for separate DEA programs for each scale regime or orientation. Moreover, because the frontier is continuously differentiable, marginal effects and elasticities are well defined along the entire surface, enabling policy-oriented counterfactuals such as capital-deepening or labor-adjustment scenarios.

Dynamic interpretation further enhances the model's explanatory capacity. Changes in the estimated frontier across years correspond to technological progress, while movement of individual firms toward the frontier captures managerial improvement or reallocation efficiency. This decomposition parallels the logic of the Malmquist productivity index but is embedded in a stochastic, differentiable framework.

Taken together, these results show that the log-log quantile frontier provides a smooth probabilistic benchmarking layer inspired by DEA, connecting efficiency to firm outcomes, supporting dynamic decomposition, and retaining economically meaningful shape restrictions in a statistically tractable form. All computations and graphical visualizations were performed using the R programming language, ensuring full reproducibility of the estimation and validation procedures.

5.3 *Relations to previous work*

The findings of this study extend and complement several major strands of research in efficiency and frontier analysis. Earlier studies [13, 14] advanced quantile-based stochastic frontier models, emphasizing heterogeneity and robustness of estimation. However, these works primarily addressed estimation strategies and did not jointly emphasize economically interpretable shape restrictions in a log-log quantile benchmark, calibration/coverage diagnostics, and explicit benchmarking comparison to deterministic DEA within one coherent framework. Our contribution is to formalize the constrained estimation and its interpretation in a way that is consistent with the probabilistic nature of quantile frontiers, rather than to claim a strict DEA-envelope equivalence.

Relative to classical DEA [2, 3], our approach introduces a DEA-inspired probabilistic benchmark that retains key production-theoretic restrictions while allowing stochastic inference and smooth differentiability. In contrast to Stochastic Frontier Analysis (SFA) [1] and developed in the modern form [11], the log-log quantile model avoids strong distributional assumptions regarding inefficiency and noise. Instead, it estimates conditional quantiles of output via convex optimization under asymmetric loss, which naturally yields a stochastic frontier without specifying parametric error structures.

The model also differs conceptually from partial-frontier methods proposed in the works [4, 5]. Consistent with recent cautions in the literature, we treat τ as a coverage/robustness parameter and do not equate quantile frontiers with partial (order-) frontiers without additional restrictive assumptions. Accordingly, increasing τ yields a more conservative near-frontier benchmark, but the object remains probabilistic rather than an envelopment surface.

By integrating these perspectives, the study unifies deterministic convex optimization and stochastic inference within a single log-log framework. The resulting benchmark complements DEA by adding smoothness, calibration tools, and inferential structure while keeping economically meaningful shape discipline within the chosen parametric class.

Several extensions naturally follow from this positioning: the constrained quantile benchmarking logic can be transferred to multi-output/network technologies and to richer dynamic panel specifications, provided that economically meaningful shape restrictions and calibration diagnostics are preserved [15, 16, 34, 37]. We discuss these extensions together with practical limitations and implementation considerations in Section 6.

6. Conclusions

This paper develops a DEA-inspired smooth stochastic benchmarking framework based on a constrained log-log quantile production frontier. On the theory side, we provide a coherent foundation for constrained quantile regression under economically interpretable inequality restrictions, including identifiability and large-sample behavior, and we clarify the probabilistic (coverage) nature of quantile frontiers. Within the Cobb-Douglas class, the log-log form yields a continuously differentiable benchmark with elasticity-based interpretation and a transparent scale index, and it supports a dynamic extension that decomposes performance into frontier shift (technical change) and relative-position change (catch-up) in a probabilistic setting. Empirically, using SPARK-Interfax panel data on 1,035 Russian manufacturing firms over 2019-2023, the proposed efficiency measures are economically meaningful: they are significantly associated with profitability and growth, while internal diagnostics (pinball loss and coverage) support calibration. Benchmarking against DEA and a parametric SFA model, together with τ-sensitivity/rank-stability checks and simple runtime micro-benchmarks, further supports the practical robustness and scalability of the implementation.

Limitations and future research. The framework is designed to complement deterministic envelopment: while DEA delivers an envelopment-based benchmark, high-quantile frontiers yield a calibrated probabilistic benchmark that is smooth and often less sensitive to noise in heterogeneous firm data. At the same time, the current empirical implementation is limited to a single-output Cobb-Douglas specification with a particular set of shape restrictions, and—as in all frontier methods (DEA, SFA, and quantile frontiers)—input endogeneity and measurement error remain important challenges. Future work will focus on two practical extensions: broader multi-sector replications using harmonized firm databases to assess portability of calibration and the TE-performance link, and richer dynamic panel implementations within the same constrained log-log framework. A closely related extension is the construction of dynamic efficiency and resilience indicators under uncertainty, together with conditions ensuring uniqueness of fitted frontier values.

Acknowledgment

This study was supported by the Russian Science Foundation (RSF), project No. 25-28-00731 “Technological leadership and digital technologies as key factors for the development of Russian firms in the context of economic instability: analysis and modeling”. https://rscf.ru/project/25-28-00731/.

Conflict of interest

The authors declare no competing financial interest.

References

[1] Aigner D, Lovell CAK, Schmidt P. Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*. 1977; 6(1): 21–37. Available from: https://doi.org/10.1016/0304-4076(77)90052-5.

[2] Charnes A, Cooper WW, Rhodes E. Measuring the efficiency of decision-making units. *European Journal of Operational Research*. 1979; 3(4): 339. Available from: https://doi.org/10.1016/0377-2217(79)90229-7.

[3] Banker RD, Charnes A, Cooper WW. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*. 1984; 30(9): 1078–1092. Available from: https://doi.org/10.1287/mnsc.30.9.1078.

[4] Cazals C, Florens JP, Simar L. Nonparametric frontier estimation: a robust approach. *Journal of Econometrics*. 2002; 106(1): 1–25. Available from: https://doi.org/10.1016/S0304-4076(01)00080-X.

[5] Daraio C, Simar L. *Advanced Robust and Nonparametric Methods in Efficiency Analysis*. New York, NY: Springer; 2007. Available from: https://doi.org/10.1007/978-0-387-35231-2.

[6] Zarrin M, Brunner JO. Analyzing the accuracy of variable returns to scale data envelopment analysis models. *European Journal of Operational Research*. 2023; 308(3): 1286–1301. Available from: https://doi.org/10.1016/j.ejor.2022.12.015.

[7] Mergoni A, Emrouznejad A, De Witte K. Fifty years of data envelopment analysis. *European Journal of Operational Research*. 2025; 326(3): 389–412. Available from: https://doi.org/10.1016/j.ejor.2024.12.049.

[8] Jung S, Son J, Kim C, Chung K. Efficiency measurement using data envelopment analysis (DEA) in public healthcare: research trends from 2017 to 2022. *Processes*. 2023; 11(3): 811. Available from: https://doi.org/10.3390/pr11030811.

[9] Holý V. Ranking-based second stage in data envelopment analysis: an application to research efficiency in higher education. *Operations Research Perspectives*. 2024; 12: 100306. Available from: https://doi.org/10.1016/j.orp.2024.100306.

[10] Färe R, Grosskopf S, Whittaker G. Network DEA. In: *Modeling Data Irregularities and Structural Complexities in Data Envelopment Analysis*. Boston, MA: Springer. 2007. p.209–240. Available from: https://doi.org/10.1007/978-0-387-71607-7_12.

[11] Tsionas MG, Kumbhakar SC. Stochastic frontier models with time-varying conditional variances. *European Journal of Operational Research*. 2021; 292(3): 1115–1132. Available from: https://doi.org/10.1016/j.ejor.2020.11.008.

[12] Tsionas MG, Assaf AG, Andrikopoulos A. Quantile stochastic frontier models with endogeneity. *Economics Letters*. 2020; 188: 108964. Available from: https://doi.org/10.1016/j.econlet.2020.108964.

[13] Fusco E, Benedetti R, Vidoli F. Stochastic frontier estimation through parametric modelling of quantile regression coefficients. *Empirical Economics*. 2022; 64(2): 869–896. Available from: https://doi.org/10.1007/s00181-022-02273-x.

[14] Papadopoulos A, Parmeter CF. Quantile methods for stochastic frontier analysis. *Foundations and Trends in Econometrics*. 2022; 12(1): 1–120. Available from: https://doi.org/10.1561/0800000042.

[15] Zheng P, Worku N, Bannick M, Dieleman JL, Weaver M, Murray CJL, et al. Robust nonparametric stochastic frontier analysis. *SSRN*. 2024. Available from: https://doi.org/10.2139/ssrn.4791808.

[16] Lee J, Han D, Park J, Choi T. Variational Bayes methods for Bayesian quantile stochastic frontier models. *Journal of the Korean Data and Information Science Society*. 2024; 35(2): 239–257. Available from: https://doi.org/10.7465/jkdi.2024.35.2.239.

[17] Wei Z, Choy STB, Wang T, Zhu X. Bayesian stochastic frontier models under the skew-normal half-normal settings. *Journal of Productivity Analysis*. 2025; 64(1): 81–91. Available from: https://doi.org/10.1007/s11123-025-00757-3.

[18] Kuosmanen T. Stochastic semi-nonparametric frontier estimation of electricity distribution networks: application of the StoNED method in the Finnish regulatory model. *Energy Economics*. 2012; 34(6): 2189–2199. Available from: https://doi.org/10.1016/j.eneco.2012.03.005.

[19] Kuosmanen T, Johnson A. Modeling joint production of multiple outputs in StoNED: directional distance function approach. *European Journal of Operational Research*. 2017; 262(2): 792–801. Available from: https://doi.org/10.1016/j.ejor.2017.04.014.

[20] Koenker R, Bassett G. Regression quantiles. *Econometrica*. 1978; 46(1): 33. Available from: https://doi.org/10.2307/1913643.

[21] Jradi S, Parmeter CF, Ruggiero J. Quantile estimation of stochastic frontiers with the normal-exponential specification. *European Journal of Operational Research*. 2021; 295(2): 475–483. Available from: https://doi.org/10.1016/j.ejor.2021.03.002.

[22] Dai S, Kuosmanen T, Zhou X. Partial frontiers are not quantiles. *arXiv:2205.11885*. 2022. Available from: https://doi.org/10.48550/arXiv.2205.11885.

[23] Dai S, Kuosmanen T, Zhou X. Generalized quantile and expectile properties for shape-constrained nonparametric estimation. *European Journal of Operational Research*. 2023; 310(2): 914–927. Available from: https://doi.org/10.1016/j.ejor.2023.04.004.

[24] Irfan M, Zunaira F, Javed M, Shongwe SC, Bhatti SH, Hussain MA. Advancements in population mean estimation for circular systematic sampling. *Contemporary Mathematics*. 2025; 6(4): 5053–5072. Available from: https://doi.org/10.37256/cm.6420257140.

[25] Chernozhukov V, Fernández-Val I, Melly B. Fast algorithms for the quantile regression process. *Empirical Economics*. 2022; 62(1): 7–33. Available from: https://doi.org/10.1007/s00181-020-01898-0.

[26] Tsionas M, Parmeter CF, Zelenyuk V. Bayesian artificial neural networks for frontier efficiency analysis. *Journal of Econometrics*. 2023; 236(2): 105491. Available from: https://doi.org/10.1016/j.jeconom.2023.105491.

[27] Chen L, Xie X, Yao Y, Huang W, Luo G. A hybrid data envelopment analysis-random forest methodology for evaluating green innovation efficiency in an asymmetric environment. *Symmetry*. 2024; 16(8): 960. Available from: https://doi.org/10.3390/sym16080960.

[28] Simar L, Wilson PW. Sensitivity analysis of efficiency scores: how to bootstrap in nonparametric frontier models. *Management Science*. 1998; 44(1): 49–61. Available from: https://doi.org/10.1287/mnsc.44.1.49.

[29] Simar L, Wilson PW. Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*. 2007; 136(1): 31–64. Available from: https://doi.org/10.1016/j.jeconom.2005.07.009.

[30] Lv W, Zhou Z, Huang H. The measurement of undesirable output based-on DEA in E&E: models development and empirical analysis. *Mathematical and Computer Modelling*. 2013; 58(5–6): 907–912. Available from: https://doi.org/10.1016/j.mcm.2013.06.007.

[31] Bondell HD, Reich BJ, Wang H. Noncrossing quantile regression curve estimation. *Biometrika*. 2010; 97(4): 825–838. Available from: https://doi.org/10.1093/biomet/asq048.

[32] Jiang R, Yu K. No-crossing single-index quantile regression curve estimation. *Journal of Business and Economic Statistics*. 2023; 41(2): 309–320. Available from: https://doi.org/10.1080/07350015.2021.2013245.

[33] Dai S, Fang YH, Lee CY, Kuosmanen T. pyStoNED: a Python package for convex regression and frontier estimation. *Journal of Statistical Software*. 2024; 111(6): 1–43. Available from: https://doi.org/10.18637/jss.v111.i06.

[34] Tone K, Tsutsui M. Dynamic DEA with network structure: a slacks-based measure approach. *Omega*. 2014; 42(1): 124–131. Available from: https://doi.org/10.1016/j.omega.2013.04.002.

[35] Färe R, Grosskopf S, Lovell CAK. *The Measurement of Efficiency of Production*. Netherlands: Springer; 1985. Available from: https://doi.org/10.1007/978-94-015-7721-2.

[36] Rani TR, Al Shibli A, Siraj M, Srimal W, Al Bakri NZS, Radhika TSL. ML-based approach to predict carotid arterial blood flow dynamics. *Contemporary Mathematics*. 2024; 5(3): 2858–2876. Available from: https://doi.org/10.37256/cm.5320243224.

[37] Tsionas MG, Tzeremes NG. Eco-efficiency estimation with quantile stochastic frontiers: evidence from the United States. *Journal of Environmental Management*. 2022; 320: 115876. Available from: https://doi.org/10.1016/j.jenvman.2022.115876.