

UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Paid for Size, Not Hardship: Diagnosing and correcting implicit structural compensation in Swedish electricity distribution benchmarking

Lukas Lindquist & Magne Björneholm

Supervisors: Johan Stennek (Department of Economics) & Erik Lundin (Institute of Industrial Economics)

Master's thesis in Economics, 30 hec

Spring 2026

Graduate School, School of Business, Economics and Law, University of Gothenburg, Sweden

Abstract

Sweden's energy regulator uses DEA to benchmark 144 electricity distribution operators against five outputs. One of these, network stations (NS), is intended to compensate firms for spatially dispersed networks. We show this compensation is systematically misallocated. At the aggregate level, NS compensation tracks density; at the firm level, a subset of predominantly urban operators, including the three largest DSOs, receive NS credit that their density does not predict. Routing density through StoNEZD's parametric z-channel corrects efficiency ratings for over-compensated firms and produces a statistically significant shift. Moreover, the current specification's exposure to the curse of dimensionality is far narrower than its five-output count implies. The effective rank of the output cloud is approximately 1.12, a single scale-axis with NS as the lone orthogonal direction. The curse of dimensionality therefore reduces to the curse of NS.

AI disclosure

AI (Claude Code in Visual Studio Code) has been used in the thesis in the following ways. Since the StoNED estimation framework is exclusive to Python, in which neither author has extensive knowledge, Claude Code has been used to create the necessary code scaffolding for us to then carry out the actual estimations ourselves. This includes data loading, pipeline configuration, path directions, and dataframe configuration: the mechanical, boring but necessary coding that would otherwise hamper our estimations significantly.

Moreover, AI has been used as a sparring partner for simplifying difficult concepts in the literature (of which there have been many...). But in the end, all creative and analytical decisions (writing, research problem, identification strategy, estimation framework, interpretation) were made and defended by the authors. Lastly, Claude Code has been used to migrate the thesis from Google Docs into LaTeX with proper syntax and to create figures through `matplotlib`, and Grammarly has been used as a proofreading aid for spelling, grammar, and clunky phrasing.

Contents

1 Introduction **4**

1.1 Rising network tariffs and revenue-cap regulation 4

1.2 Yardstick competition, heterogeneity, and benchmarking 4

1.3 Network stations, dimensionality, and the parametric alternative 5

1.4 Research gap, questions, and contributions 8

2 Theory and Literature Review **10**

2.1 Yardstick Competition and Firm Heterogeneity 10

2.2 Compensating Residual Heterogeneity in Practice 12

2.2.1 Norway: Two-Stage Regression 12

2.2.2 Finland: Joint Parametric Estimation 13

2.3 Dimensionality and Overfitting in Frontier Estimation 14

2.3.1 The Common Geometry of DEA and CNLS 14

2.3.2 How Overfitting Compounds with Dimensionality 15

2.3.3 Effective Rank and Intrinsic Dimension 16

2.3.4 DEA: Weight Optimisation 17

2.3.5 CNLS: Subgradient Overfitting 17

2.3.6 Remedies: Bounding Weights and Slopes 18

3 Data **20**

3.1 Regulatory benchmarking data 20

3.2 Output cloud geometry 21

3.3 Concession geography and fragmentation 22

3.4 Environmental z-variables 23

3.4.1 Exposure-aligned weighting principle 23

3.4.2 Population density 24

3.4.3 Snow 25

3.4.4 Forest 25

4 Methodology **27**

4.1 Specification context 27

4.2 Data Envelopment Analysis (DEA) 27

4.3 Stochastic nonparametric envelopment of data (StoNED) 28

4.3.1 StoNED 28

4.3.2 StoNEZD 30

4.4 Estimation design 30

4.4.1 Specifications 30

4.5 Diagnostics 31

4.5.1 Diagnosing NS as a compensation instrument (RQ1) 31

4.5.2 Evaluating the parametric z-channel (RQ2) . . . 33

5 Results **36**
5.1 Network Stations as Implicit Structural Compensation (RQ1) . . . 36
5.1.1 How DEA and StoNED distribute output weight . . . 36
5.1.2 Network stations as the dominant flexibility driver . . . 37
5.1.3 NS dependence, rurality, and peer structure . . . 38
5.1.4 Firm-level NS–density alignment . . . 39
5.2 The Parametric z-Channel as Explicit Compensation (RQ2) . . . 41
5.2.1 Delta-hat inference . . . 41
5.2.2 Cohort falsification . . . 42
5.2.3 MAUP sensitivity . . . 43

6 Discussion **45**
6.1 NS reliance and the limits of implicit environmental compensation . . . 45
6.2 The curse of dimensionality is a curse of NS . . . 45
6.3 Can StoNED deliver what DEA cannot? . . . 46
6.4 Implications for RP5 . . . 47

7 Conclusion **48**

A Supplementary Material **54**
A.1 Best practice versus average practice frontier . . . 54
A.2 Concession dispersion . . . 55
A.3 Output cloud orthogonality . . . 56
A.4 Revision report . . . 57

1 Introduction

1.1 Rising network tariffs and revenue-cap regulation

Between 2010 and 2020, real electricity network tariffs in Sweden rose by 37 percent (Lundin and Söderberg, 2022). Whether that increase reflects warranted investment or a regulatory framework that has not held prices to efficient levels is an open question. While cross-country comparisons must be interpreted with caution given differences in regulatory design and market conditions, the increase places Sweden as an outlier, exceeding those observed in Finland[1] (29 percent), Denmark (3 percent), and Norway (0 percent) over the same time period (Lundin and Söderberg, 2022). Observable cost drivers, particularly wind power expansion, partially explain Sweden's outlier position, but regression analysis shows that measurable factors cannot fully account for the price trajectory, and improved supply quality cannot explain the remainder since interruption durations remained roughly constant. The price increases were particularly concentrated among the three largest DSOs (Vattenfall, E.ON, and Ellevio, which together serve roughly half of all customers), where prices rose by 52 percent compared to 26 percent for smaller firms (Lundin and Söderberg, 2022). Assessing whether the regulatory framework is implicated in the unexplained price growth requires understanding what that framework is designed to do.

Electricity distribution is a natural monopoly: the fixed costs of building and operating network infrastructure are large relative to the marginal cost of serving additional customers, making parallel networks economically wasteful (Joskow, 2007). Each Swedish distribution system operator (DSO) holds an exclusive network concession granting it the sole right to distribute electricity within a defined geographic area, meaning that consumers are therefore unable to choose their provider within this market. In a competitive market, rival firms would pressure each other toward cost-efficient operation; a monopolist faces no such pressure and, absent intervention, could operate above efficient cost levels (Laffont and Tirole, 1993). Sweden disciplines this through ex-ante revenue cap regulation (Joskow, 2008) where the Energy Markets Inspectorate (Energimarknadsinspektionen, hereafter Ei) sets maximum allowable revenue for each DSO over a four-year regulatory period. Under a binding revenue cap, every SEK of cost reduction below the allowed level accrues directly to profit for the firm, giving the firm incentive to operate cost efficiently.

1.2 Yardstick competition, heterogeneity, and benchmarking

Setting the revenue cap to efficient levels, however, requires the regulator to determine what an efficient cost level is for each firm, and this is fundamentally a problem of information asymmetry:

[1] Following two severe storms in 2011 that caused extended outages for roughly 17 percent of households, the Electricity Market Act was amended to establish maximum permissible outage durations, prompting extensive replacement of overhead lines with underground cables (Lundin and Söderberg, 2022). Finland's price increase is therefore motivated and reflects mandated quality investment.

the firm knows its own cost structure, operating conditions, and true efficiency potential far better than the regulator does. A regulator that simply reimburses reported costs leaves the firm with no reason to economise, since profit no longer depends on the cost choice, and even theoretically optimal designed regulatory contracts cannot fully remove the firm's information rent (Laffont and Tirole, 1993). Yardstick competition (Shleifer, 1985) is the canonical theoretical response to this problem. Rather than trying to determine each firm's efficient cost directly, the regulator benchmarks each firm against its peers, tying a firm's allowed revenue not to its own reported costs but to what comparable firms achieve, leaving firms with no instrument to influence their own allowed revenues and the optimal response is to operate cost efficiently; the comparison reproduces the cost level a fully informed regulator would set, inducing competition in otherwise isolated monopolies.

Heterogeneity breaks this construction, producing the efficient outcome only if the regulator observes the heterogeneous characteristics and adjusts the yardstick accordingly (Shleifer, 1985). Real DSOs are not identical entities: they operate in environments that differ in structural elements outside their control (customer density, terrain, climate, vegetation) that affect cost. If the yardstick is not adjusted for these differences, the comparison is distorted in two directions which carry real welfare consequences. Penalising a genuinely managerially efficient firm for operating in costlier environments imposes an efficiency requirement it cannot meet through better management; the firm can comply only by inefficient cost-cutting, deferring maintenance and network renewal which undermines supply quality in the short term and efficient investment incentives in the long term. At the other end, a managerially inefficient firm operating in a favourable environment faces weak incentives to operate cost-efficiently, preserves allocative inefficiency, and transfers rent from consumers to firms through above-efficient prices.

Benchmarking is the empirical implementation of Shleifer (1985)'s yardstick competition. Among frontier methods regulators use to construct the efficient-cost norm from observed practice, Data Envelopment Analysis (DEA), introduced by Charnes et al. (1978), is the most widely adopted (Arocena et al., 2025; Bogetoft, 1997) and is the method Ei uses. DEA is nonparametric: rather than assuming a functional form for the cost function, it derives the cost frontier directly from the firms that deliver a given level of output at the lowest cost and benchmarks every other firm against this best practice. Its regulatory appeal is that it assumes almost no structure on the data.

1.3 Network stations, dimensionality, and the parametric alternative

As of regulatory period 4 (RP4), Ei's benchmarking specification rests on five outputs which can be separated into two strands. Four are service outputs each measuring what the firm actively delivers to its customers. The fifth, the number of network stations (NS), is different in kind and is best read as a measure of the physical network extent and is the instrument through which the specification compensates firms for structural heterogeneity in how concentrated or dispersed

their networks are. That a count of transformer stations can stand in for spatial dispersion at all rests on a quiet assumption: NS needs to carry variation along an orthogonal axis that the four service outputs fail to capture. Chapter 3 and Section 2.3 develop this geometry and its consequences.

Placing NS in the output vector, however, asks one variable to do two different jobs. On one hand, a network station is a genuine piece of capital with building and maintaining one costing money, so the count of NS is a genuine piece of the production technology. On the other hand, NS is asked to carry a structural correction and compensate firms operating in less favourable concessions in terms of spatial dispersion that they themselves did not choose. Within the general benchmarking literature, the two readings call for opposite treatment: the production technology should shape the frontier while a structural factor should leave the frontier intact and only reposition a firm relative to it (Daraio et al., 2018). Yet, NS in the output vector conflates these two and routes both through the same channel and we cannot distinguish the magnitude of each. Chapter 2 formalises the conflation. However, regulators are not restricted to operationalising structural compensation through proxies in the output vector, and several options for adjusting a firm's position relative to the frontier while leaving frontier shape intact exist.

The second option runs through a different strand of estimation. One option stays in the DEA framework and addresses heterogeneity through second-stage regression of first-stage efficiency scores on structural variables, but the approach suffers from statistical problems (Johnson and Kuosmanen, 2012; Simar and Wilson, 2007). The alternative this thesis investigates is found in a different frontier estimation method. Stochastic Nonparametric Envelopment of Data (StoNED), introduced by Kuosmanen and Kortelainen (2012), combines nonparametric frontier estimation with a stochastic decomposition of each firm's distance from the frontier into genuine inefficiency and random noise — a stochastic treatment DEA lacks. The frontier underlying StoNED is estimated via convex nonparametric least squares (CNLS), is an average-practice surface fitted through the cloud of firms which contrasts the best-practice envelopment in DEA (see Figure A.1 in Appendix A.1 for illustration). Despite the stochastic treatment inherent to the method, StoNED on its own is not the method of interest, rather an extension of it. Stochastic Non-smooth Envelopment of z-variable Data (StoNEZD), developed by Johnson and Kuosmanen (2011) and the basis of the Finnish energy regulator's cost benchmarking, lets a structural variable enter the model through a separate parametric channel — the z-channel — rather than through the output vector.

Before developing that alternative any further, a second complication with the frontier estimation must be discussed. The conflation of placing NS in the output vector is at heart conceptual, but sitting on top of it is a concrete finite-sample problem. Although DEA and CNLS produce different artefacts (a best-practice frontier that envelopes observations versus an average-practice frontier through them), they estimate the same underlying geometric object: a piecewise-linear convex surface assembled from firm-specific supporting hyperplanes whose orientation is disciplined by neighbouring firms (Charnes et al., 1978; Kuosmanen, 2008; Kuosmanen and Johnson, 2010). How these firm-specific hyperplanes are estimated differs across the two methods, but the distinction

is not essential for the intuition; the formal development is deferred to Section 2.3.

Picture a thief slipping away from a robbery through a crowd. In the thick of the crowd, bodies press in from every side and there is no direction to break for; but if the thief happens to be at the outer edge of the crowd, only a person or two stands between the thief and freedom. A firm in DEA and CNLS is disciplined in much the same way. When many firms closely resemble it in scale and output mix, they restrict its position from all sides and leave the model little room to rate it more efficiently than it truly is. When few firms resemble it, there is little room for discipline and the model will happily exploit any available slack to find a direction that makes it look efficient. What governs the thinning of disciplinary neighbours, though, is not the nominal output count but the number of directions the output cloud genuinely occupies. Chapter 3 estimates the number of effective dimensions for the Swedish specification at just above one: a single scale axis of the four service outputs, with NS the lone direction the cloud adds to it. In our analogy, any escape direction that enables our thief to make a run from the thick of the crowd will have a big glowing sign with NS on it. The variable intended to make only minor corrections for spatial dispersion may, in fact, be the one that shapes the frontier the most. Section 2.3 makes this argument precise and gives it its formal name, the curse of dimensionality.

Taken together, the two problems (how to compensate structural heterogeneity, and the propensity of nonparametric frontiers to overfit) point in the same direction: route the structural adjustment through a different placement, and this is what StoNEZD provides, entering parametrically through a single global coefficient that only adjusts the firm's relative position to the frontier and not its shape. Because the variable does not enter the frontier, it does not add to dimensionality, and the boundary slack of the previous paragraph cannot reach it. Another advantage is that fitted cost splits cleanly into a production component and a structural component, and the structural component is subject to conventional statistical inference.

However, the advantages are not unconditional. The z-channel requires that each structural dimension is reduced to a single scalar per firm, aggregated from gridded geographical data covering the firm's operating area. Any reasonable aggregation rule for firms' effective exposure essentially yields the same scalars for firms with compact, continuous operating areas; but for firms with large or geographically dispersed operating areas across the country — Vattenfall, E.ON, and Ellevio above all — it becomes an explicit modelling choice since defensible alternatives for effective exposure may yield materially different scalars. This is the Modifiable Areal Unit Problem (Fotheringham and Wong, 1991; Openshaw, 1984), its implication being that z-variable construction is part of the model and not purely data preparation, and any sensitivity propagates downstream to every object the z-variable touches. The z-channel, then, does not so much abolish the difficulty of compensating dispersion as relocate it: from a flexibility problem inside the frontier to a measurement problem in the z-vector, and it relocates it onto precisely the firms whose compensation was least clear-cut to begin with: Vattenfall, E.ON, and Ellevio, hereafter the *Big Three*. Whether density can be routed through the z-channel cleanly enough for those firms is therefore an empirical question, and one this thesis investigates.

1.4 Research gap, questions, and contributions

The literature on heterogenity treatment in Nordic regulatory benchmarking has compared the two-stage approach (Norway) against the parametric treatment (Finland) and found the latter dominant in statistical properties (Johnson and Kuosmanen, 2011) and simulation studies (Cordero et al., 2020; Johnson and Kuosmanen, 2011; Nieswand and Seifert, 2018), with the applied literature documenting systematic firm-level miscompensation in the Norwegian framework (Bjørndal et al., 2016). The third option of embedding structural proxies in the output vector (Sweden) sits outside the comparison. Whether this third approach delivers structural compensation to the firms whose operating environment warrant it, given the intrinsic overfitting pathology developed in Section 2.3 and Chapter 3, is an empirical question that has not been asked. The answer also carries more weight when considering recent regulatory developments. Historically, the efficiency requirements from benchmarking has had little practial force: revenue frame has in most cases been non-binding, the efficiency requirement has been applied to operating expenditures, and benchmarking was one of several instruments disciplining overinvestments. The forthcoming fifth regulatory period (RP5) changes the equation: revenue frames are expected to tighten and bind after a change in indexation of how the capital base is valued; the requirement will now apply to total expenditures, enlarging the cost base it bears on; and the parallel disciplinary instruments are folded into the TOTEX benchmarking (Energimarknadsinspektionen, 2024). Benchmarking thereby becomes the primary instrument disciplining firm cost, so any systematic misallocation in how it compensates structural heterogeneity now propagates directly into the binding frame, with no parallel instrument left to offset it.

The question therefore becomes sharper at the firm level. Aggregate diagnostics may show that NS-based compensation tracks density on average but leave open whether the firms collecting most of that compensation are the firms whose conditions warrant it. If the answer is no, the natural next question is whether StoNEZD's parametric shift relocates compensation in the direction the diagnostic predicts. Two questions follow, we ask:

RQ1. Within the current regulatory specification, does the structural compensation that NS provides reach the firms whose density profile warrant it?

RQ2. When structural compensation is instead routed through StoNEZD's parametric z-channel and NS is dropped from the output vector, do the firms RQ1 identifies as miscompensated relocate in the direction the critique predicts, and is the relocation robust to how the z-vector is constructed for firms with geographically dispersed concession areas?

Contributions and preliminary findings. This thesis makes three contributions. First, it documents a systematic misalignment in how Ei's current specification delivers structural compensation. Network stations (NS), included in the output vector to proxy operating-environment difficulty (population density), disproportionately benefit the three largest DSOs; not because they serve unusually harsh terrain, but because their networks span multiple density zones. We show this through a sequence of diagnostics that move from aggregate to firm-level: output-weight

distributions and leave-one-variable-out tests establish that NS dominates frontier flexibility in both DEA and StoNED, while a fractional regression of NS weight on population density confirms that the extra flexibility moves to firms whose density profiles do not warrant it. The misalignment is robust to additional environmental controls and constructions of density.

Second, we show that routing density through StoNEZD's parametric z-channel corrects the most prominent part of this misalignment. Over-compensated firms — headlined by the Big three — see both their fitted cost and efficiency rating decrease when density enters the z-channel rather than the output vector, and the correction is stable across four density measures and two environmental-control configurations. Crucially, the z-channel makes the size and statistical significance of the density adjustment transparent and testable; a property absent from the firm-specific NS multipliers that DEA produces, which are not subject to inference.

Third, we show that the Swedish specification's exposure to the curse of dimensionality is far narrower than its five-output count implies and its implication holds true for both frontier methods. DEA and CNLS estimate the same geometric object — a piecewise-linear convex hypersurface built from firm-specific supporting hyperplanes whose orientation is disciplined only by neighbouring firms — and we argue that this shared geometry carries a common geometric overfitting pathology that compounds in intrinsic rather than nominal terms. On a nominal reading the five-output count would flag the specification for the curse of dimensionality, but computing the effective rank of the output cloud places the effective dimension just above one: a single scale axis with NS as the lone orthogonal direction. The dimensionality concern and the escape directions it opens therefore reduces almost entirely to one variable: in the Swedish specification, the curse of dimensionality is a curse of NS.

2 Theory and Literature Review

2.1 Yardstick Competition and Firm Heterogeneity

Yardstick competition and benchmarking by extension presupposes that the regulator can compare firms on common terms. Shleifer (1985) provides the canonical theoretical implementation. Consider $N \geq 2$ regulated firms, each a monopoly serving a separate market with downward-sloping demand $q(p)$. Each firm starts with constant marginal cost c_0 and can decrease it to c_i by spending $R(c_i)$ on cost reduction, where $R(c_0) = 0$, $R'(c) < 0$, and $R''(c) > 0$. The regulator faces an information asymmetry problem: it observes realised cost c_i but not the cost-reduction technology $R(\cdot)$. Define firm i's profit as

$$V_i = (p_i - c_i)\, q(p_i) + T_i - R(c_i), \tag{2.1}$$

where p_i is the regulated price and T_i a lump-sum transfer.

A regulator who maximises total surplus subject to firms' break-even condition $V_i \geq 0$ obtains the familiar first-order conditions

$$p^* = c^*, \qquad -R'(c^*) = q(p^*), \qquad T^* = R(c^*), \tag{2.2}$$

where regulated price equals marginal cost, the transfer just covers cost-reduction, and the marginal cost of reduction equals its marginal benefit.

Traditional cost-of-service regulation (Joskow, 2007) can not reach this socially efficient optimum: setting $p_i = c_i$ and $T_i = R(c_i)$ for whatever cost the firm reports makes $V_i \equiv 0$ and firm profits independent of its cost choice, so the manager sets $c_i = c_0$ and reduces nothing.

Yardstick competition severs the link between firm profit and its cost choice. The intuition is simple: tying firm i's transfer to peer comparison leaves the firms with no instrument to influence its own price, and the optimal response is to minimise expenditures to socially optimal levels in aggregate.

For each firm i, define *shadow-firm* from the average cost of its rivals

$$\bar{c}_i = \frac{1}{N-1} \sum_{j \neq i} c_j, \qquad \bar{R}_i = \frac{1}{N-1} \sum_{j \neq i} R(c_j), \tag{2.3}$$

and let the regulator commit to price and transfer rule $p_i = \bar{c}_i$ and $T_i = \bar{R}_i$. Substituting this into firms profit function (2.1) yields

$$V_i = (\bar{c}_i - c_i)\, q(\bar{c}_i) + \bar{R}_i - R(c_i) \tag{2.4}$$

and maximising (2.4) over c_i yields the first-order condition

$$-R'(c_i) = q(\bar{c}_i) \tag{2.5}$$

which equals the regulator's cost-reduction condition in (2.2). In the homogeneous firm case, the firms choose $c_i = c^*$ and the unique Nash equilibrium of the simultaneous game coincides with the social optimum without the regulator observing $R(\cdot)$.

However, heterogeneity breaks this construction. If firms differ in observable, exogenous characteristic vector z_i, averaging peers' costs is no longer informative for firm i's benchmark because they operate under different structural conditions. The *reduced-form regulation* of Shleifer (1985) extends the game to include heterogeneity between firms. With heterogeneous firms, indexing the problem faced by the regulator in (2.2) by z gives the heterogeneous analogue

$$-R_c(c, z) = q\big(c(z)\big), \tag{2.6}$$

of the cost-reduction condition. Taylor expanding (2.6) around some reference point makes the first-best cost approximately linear in characteristics $c \approx a + bz$ which the regulator estimates through cross firm regression of realised costs on z

$$\hat{c}_i = \hat{a} + \hat{b}\, z_i. \tag{2.7}$$

The regulator then commits to price rule

$$p_i = \hat{c}_i, \tag{2.8}$$

and replaces shadow firm (2.3) with $\hat{c}_i$. However, the extension only restores the first best under a strong condition. If we decompose each firm's true first-best cost $c^*(z_i)$ into a part explained by observed characteristics and an unexplained residual

$$c^*(z_i) = \underbrace{\hat{a} + \hat{b}\, z_i}_{\hat{c}_i} + \varepsilon_i, \tag{2.9}$$

where ε_i comes from the regression in (2.7), reduced form regulation only recovers the social optimum if z contains every structural cost driver such that $R^2 = 1$ and $\varepsilon_i = 0$. If $R^2 < 1$, residual heterogeneity survives in ε_i and propagates into regulated price p_i. A firm priced at $p_i = \hat{c}_i$ earns realised rent

$$V_i = (\hat{c}_i - c_i)\, q(\hat{c}_i) + \bar{R}_i - R(c_i, z_i), \tag{2.10}$$

which equals zero when $\varepsilon_i = 0$ but carries the sign of $-\varepsilon_i$ otherwise. An under-compensated firm – one whose true cost exceeds what its observable characteristics predict, $\varepsilon_i > 0$ – receives a regulated price and transfer below its efficient cost; it participates at a loss and exits in the absence of external subsidy. A structurally advantaged firm, $\varepsilon_i < 0$, is over-compensated and extracts rents. If the regulator's task is to achieve the social optimum, it must account for

heterogeneity and construct a benchmark in which $\varepsilon_i = 0$ for all $i = 1, 2, \ldots, N$.

2.2 Compensating Residual Heterogeneity in Practice

Shleifer's reduced-form regulation reduces to a single empirical problem: construct a benchmark for which $\varepsilon_i = 0$. The regression in equation (2.7) leaves a residual whenever z misses a structural cost driver, and that residual propagates one-for-one into the regulated price. Reducing the residual to $\varepsilon_i = 0$ is easier said than done and there is no consensus on how to do it in practice. This section reviews how two Nordic regulators handle the residual: project it onto z in a separate stage (Norway), or absorb it into a parametric global shift $\delta' z$ (Finland).

2.2.1 Norway: Two-Stage Regression

The Norwegian Regulator (RME) operates a three-stage cost-norm model on roughly a hundred local distribution operators. Stage 1 estimates a CRS-DEA frontier with TOTEX as the sole cost variable and three outputs (number of subscribers, kilometres of high-voltage network, and number of network stations). Stage 2 regresses the stage-1 score by OLS on a set of environmental variables (forest cover, coast exposure, snow days, and similar geographic conditions) where each variable enters as the firm's own value minus the average value of its DEA peers, producing a firm-specific correction. Stage 3 calibrates aggregate cost to actual industry cost (Kordahl et al., 2022). The architecture follows Ray (1988): compute DEA on (x, y) alone, then project the score on z separately. The construction keeps the production possibility set free of z, answering the foundational critique of Ruggiero (1998) that imposing convexity and free disposability over a structural factor improperly restricts the technology set — a firm can not substitute snow for cables.

The approach is however not without its critique. In the seminal paper, Simar and Wilson (2007) prove that stage-1 efficiency estimates are serially correlated in finite samples which is sufficient to invalidate conventional second-stage standard errors and propose smooth bootstrap truncation. Banker et al. (2019), however, argues that the critique is overstated, showing that simple DEA + OLS dominates the bootstrap-truncated procedure on mean absolute deviation when simulated under DGPs with two-sided noise which Simar and Wilson (2007) omits. However, Johnson and Kuosmanen (2012) prove that the finite-sample bias[1] of the stage-1 DEA estimator carries over to the stage-2 regression coefficients whenever z correlates with the outputs underlying the Stage 1 efficiency scores. Since the very rationale for including z is that these are cost drivers — otherwise there would be no need to account for them — that influence and therefore correlate with firm's network infrastructure, it stands to reason that non-zero output–z correlation is the standard case and not the exception. The bias affects point estimates, not just standard errors, and it seems that no inference correction may alleviate it.

A closely related defect is the *separability* condition introduced in Daraio et al. (2018), which formalise the implicit assumption of any two-stage construction. The condition states environ-

[1] $\text{Bias}(\hat{\delta}^{\text{2-DEA}}) = -(Z'Z)^{-1}Z'\,\text{Bias}(\hat{\phi}^{DEA}(X))$.

mental factors may shift where firms sit relative to the frontier but may not shift the frontier itself; $\Psi^z = \Psi$ must hold for all z. They develop a CLT-based statistical test, reject separability on a widely cited two-stage DEA dataset, and note that the same assumption has been implicitly maintained across hundreds of subsequent applications. If separability fails, Ψ is the envelope of conditional technologies that no single firm actually faces, and the stage-1 score no longer measures distance to the appropriate boundary.

The RME's response to the critique have been post-hoc operational adjustments rather than methodological adjustments, rejecting Bjørndal et al.'s conclusion that the two-stage method "[. . .] is likely to capture multiple effects and lead to under- or overcompensation of the individual regulated firms [. . .]" Bjørndal et al., 2016, p. 712 on the grounds that none of the considered empirical alternatives are preferred but capping the stage-2 correction to 40 % as of 2023 (Kordahl et al., 2022). Moreover, the z-variable *share of high-voltage underground cable* was removed from stage 2 due to endogeneity — the share of cabling is a technology choice that firms can directly influence — which would otherwise violate the exogeneity condition required for second stage regression to be consistent (Simar et al., 2016).

2.2.2 Finland: Joint Parametric Estimation

The Finnish Energy Authority (Energiavirasto) regulates roughly ninety distribution firms under a conditional yardstick (for a full treatment, see Kuosmanen, 2018; Kuosmanen and Johnson, 2021) framework that benchmarks controllable OPEX conditional on a fixed capital stock instead of traditional OPEX-CAPEX or TOTEX yardsticks. The frontier is estimated by StoNEZD (Johnson and Kuosmanen, 2011) with the multiplicative specification and three outputs (transmitted energy weighted by voltage level, total network length, number of metering points), outage costs as an undesirable output (Kuosmanen and Johnson, 2017; Saastamoinen and Kuosmanen, 2016), and the ratio of connections to metering points (C/M) enters parametrically as z (Energiavirasto, 2023a,b). Joint parametric estimation is the explicit response to Johnson and Kuosmanen (2012) — there is no stage 1 bias propagating into a stage 2. Moreover, similar to the Norwegian case, in 2016 the Finnish regulator replaced the percentage of underground cabling with the C/M ratio as the lone z due to problems with endogeneity (Kuosmanen and Johnson, 2021).

Joint estimation of heterogeneity in benchmarking closes the two-stage bias channel at the cost of introducing a separate defect:

> [The] StoNEZD method does not allow for considering the effect of contextual variables on the shape of the boundary, since it includes the impact of the contextual variables as part of the error term of the production function. Therefore, these factors can only have influence on the distribution of the inefficiencies. This can be considered as the major theoretical shortcoming of this method, since its implementation requires verifying previously whether the separability condition holds using some statistical tools such as those suggested by Daraiao & Simar, 2018. Cordero et al., 2020, p. 2261

Three consequences follow. First, the same Daraio et al. (2018) separability burden Norway carries is also Finland's: the $\delta' z$ shift alters the inefficiency conditional density without touching the frontier shape, which is licensed only under separability. Second, $\delta' z$ is a *uniform* shift: every firm receives the same marginal compensation per unit of z, regardless of where it sits in input space. Third, because $\delta' z$ requires a firm-level scalar for each contextual factor — whether a structural ratio defined on firm aggregates (C/M) or raster/grid data on density, vegetation, or climate — the Modifiable Areal Unit Problem (MAUP) (Fotheringham and Wong, 1991; Openshaw, 1984) is introduced. The aggregation rule for firms' effective exposure is innocuous for compact, continuous concessions but becomes an explicit modelling choice for firms with large or geographically dispersed operating areas (Figure A.2 in Appendix A.2), with any sensitivity in the aggregation propagating downstream to the heterogeneity correction.

Monte Carlo evidence supports StoNEZD's dominance across the relevant comparators. Johnson and Kuosmanen (2011) show that at StoNEZD produces lower RMSD of $\hat{\boldsymbol{\delta}}$ than two-stage DEA at virtually every combination of n and $\rho(\mathbf{x}, \mathbf{z})$, with the advantage largest when $|\rho|$ is high — the setting in which two-stage DEA's first-stage bias is most severe. The dominance persists even in the deterministic $\sigma_v = 0$ scenario, where two-stage DEA's formal assumptions are satisfied. Cordero et al. (2020) further simulate under Cobb–Douglas, translog and polynomial DGPs at $n \in \{50, 200\}$ with varying input–z correlation, comparing StoNEZD, conditional DEA[2], and two-stage DEA + OLS, showing that StoNEZD achieves lower mean absolute deviation and higher correlation with true efficiency in every scenario, although its performance deteriorates somewhat when z and x are correlated. Nieswand and Seifert (2018) construct a design closer to a realistic regulatory DGP — unbalanced firm sizes, right-skewed z — and find StoNEZD competitive with latent-class SFA and systematically better than conditional DEA, noting that "[...] our results therefore endorse the Finnish regulator which uses StoNEZD for such a setting in the electricity distribution sector [...]" Nieswand and Seifert, 2018, p. 146.

2.3 Dimensionality and Overfitting in Frontier Estimation

2.3.1 The Common Geometry of DEA and CNLS

DEA and CNLS produce different artefacts – a best-practice frontier that envelopes observations versus an average-practice frontier through them – but estimate the same geometric object: A piecewise-linear convex hypersurface constructed from firm-specific supporting hyperplanes whose orientation is disciplined by its neighboring firms (Charnes et al., 1978; Kuosmanen, 2008; Kuosmanen and Johnson, 2010). The following section develops the stochastic geometric pathology common to both methods, how and why it compounds with dimensionality, and documents how it manifests differently under both methods.

Some housekeeping are in place. "Neighbour" operationalises differently across the two methods, but its implication – the disciplinary action – is the same supporting hyperplane in two algorithmic formulations. On the DEA side the peer definition and intuition is simple: peer h disciplines firm

[2]See (Daraio and Simar, 2005) and (Daraio and Simar, 2007).

i when its intensity weight in the envelopment form is strictly positive, that is, when peer h forms part of the frontier firm i is compared to. On the CNLS side the same disciplinary mechanism is given by the Afriat inequalities: under the multiplicative CRS specification used throughout, firm i's subgradient $\boldsymbol{\beta}_i$ must satisfy $\boldsymbol{\beta}_i'\mathbf{y}_i \geq \boldsymbol{\beta}_h'\mathbf{y}_i$ for every other firm h, and only the inequalities with zero slack in finite samples actively bind (Afriat, 1972; Hildreth, 1954). Kuosmanen (2008) makes the formal bridge between the two methods: CNLS solves a varying-coefficient regression in which each firm carries its own slope vector $\boldsymbol{\beta}_i$, and DEA is the special case where residuals are sign-constrained to be non-positive.

Two terms follow. The slope vector $\boldsymbol{\beta}_i$ is firm i's vector of estimated marginal costs; geometrically it is the supporting hyperplane of the frontier at $\mathbf{y}_i$, and in the convex-regression literature on consistency it is the *subgradient* of the fitted cost function. A *boundary firm* is one whose slope vector is disciplined by a few or no observations. On the DEA side this is an exterior self-evaluators – an efficient frontier firms not in any other firms peer structure and serves as isolated benchmarks (Torgersen et al., 1996) on the other part of the frontier, potentially informative about unique input-output combinations (Edvardsen et al., 2008). On the CNLS side it is a firm whose subgradient is unbounded in probability at the boundary of the data: Ghosal and Sen (2017, Lemma 5.1) in the univariate case and Liao et al. (2026, Theorem 2) for the multivariate case. With the geometry taken care of, the next question is what happens when the number of variables grows.

2.3.2 How Overfitting Compounds with Dimensionality

Increasing variable count d with sample size n carries two geometric consequences: a larger share of firms end up on the convex hull – boundary – of the data instead of at the interior, and each firm gains an additional direction in which no neighbor may discipline its hyperplane. The boundary-share point has formal backing in convex geometry and random-polytope theory: The expected convex-hull vertex count of n iid points in $\mathbb{R}^d$ as $T(P)(\log n)^{d-1}$ for polytope support and $n^{(d-1)/(d+1)}$ for smooth support, and that these laws apply in the intrinsic dimension d^* whenever the support lies in a d^*-dimensional affine subspace (Bárány, 2008; Bárány and Larman, 1988, Theorem 9.1 and eq. 6.3); the result is independent of estimation method and lives entirely in the input-output cloud. The application of convex geometry and random-polytope theory is suggestive rather than literal – observations are not iid uniform draws from a known convex body with nonempty interior – but it guides what to expect in applied settings.

Secondly, increasing variable count d with sample size n means each boundary firm gains an additional direction in which no firm disciplines its hyperplane. The geometric interpretation is symmetric across both methods and is suprisingly intuitive: the constraint can only bind where a neighbor is located, so directions – however close or far they are – of the input-output space without a neighbor are unconstrained. The important point is that these two consequences compound: more outputs simultaneously push more firms to the boundary while increasing escape-directions per firm. An analogy may be in place: think of a thief slipping away through a crowd after a robbery. In the thick of it, bodies block every direction the thief might make a

break for, and there is no way out. At the edge, only a few people stand between the thief and freedom, and the thief walks straight through a gap.

Both literatures formalise the dimensional cost as nonparametric convergence rates whose denominator structure share a common form, although the loss function across the two methods differing. On the DEA side, for smooth (twice-differentiable) frontiers, Kneip et al. (1998) establishes that the VRS estimator converges pointwise at rate $n^{-2/(p+q+1)}$ in p inputs and q outputs, while Park et al. (2010) shows that the CRS case converges at $n^{-2/(p+q)}$, effectively reducing the effective dimensions. On the side of convex regression with squared error loss over the entire domain, Han and Wellner (2016) establishes the optimal mean-square rates to $n^{-2/(d+1)}$ on domains with smooth boundaries and $n^{-4/(d+4)}$ on polytope supports, refining prior rates derived in Stone (1980). Kur et al. (2020) sharpens the general convex regression rates to the CNLS estimator, showing that it attains the minimax rate of $n^{-4/(d+4)}$ on polytope supports up to $d = 4$ but is strictly suboptimal at $n^{-2/d}$ for $d \geq 5$, which the authors credit to piecewise-linear overfitting instead of boundary geometry of the kind studied in (Korostelev et al., 1995). The two denominators, $p + q + 1$ on the DEA side and $d + 1$ on the convex-regression side, can be interpreted as qualitatively equivalent $p + q + 1 \leftrightarrow d + 1$ since it is the dimension of the convex object under estimation. This infers that under matching smoothness conditions – smooth frontier on the DEA side and a smoothly bounded domain on the CNLS side – the two rates point to the common geometry of estimating the boundary of a smooth convex body. In both cases the mechanism is the one identified above: as d grows, more firms sit on the convex hull with fewer disciplinary neighbors. However, all sets of variables are not created equally, what compounds in finite samples is not the nominal dimensionality d but the intrinsic dimensionality of the output cloud: the relevant measure for escape directions.

2.3.3 Effective Rank and Intrinsic Dimension

An output nearly collinear with existing ones changes the geometry only marginally, the boundary set barely grows and each firm gains few new escape directions. An output roughly orthogonal to those already in the specification does the opposite, opening up a fresh axis along which previously interior firms can escape. The formal handle is the effective rank of the output cloud $r_e(\Sigma) = \mathrm{tr}(\Sigma)/\|\Sigma\|_{\mathrm{op}}$ (Koltchinskii and Lounici, 2017; Vershynin, 2026) which measures the directions the output cloud genuinely occupies. The effective rank is bounded by the algebraic rank and increases roughly by one with each orthogonal addition. The treatment and consideration of effective rank within frontier estimation differs between the two methods.

On the DEA side, the literature has formalised a closely related conclusion through a different apparatus: Wilson (2018) argues that estimators in nonparametric frontiers suffer the curse of dimensionality in intrinsic rather than nominal terms, introducing two diagnostics (*effective parametric sample size* and leading eigenvalue share $R_x = \lambda_{x,1}/\sum_j \lambda_{x,j}$) for collinearity analysis, while Adler and Golany (2001, 2002), Adler and Yazhemsky (2010), and Jenkins and Anderson (2003) reach the same intuition through PCA-DEA and a variable reduction approach respectively. All strands identify the same symptom: too many inputs and outputs relative to firms produce

too many efficient units and degrades discriminatory power.

On the CNLS side, to our knowledge, the literature has not formalised the link between effective rank and the subgradient pathology: Liao et al. (2026, Theorem 2) treats the boundary point as a geometric corner of the support without engaging with the output cloud's eigenvalue spectrum. The refinement this section adopts is that the boundary-share and escape-direction compounding both scale in $r_e(\Sigma)$ rather than in nominal d: correlated additions to the output vector inflate the pathology less than orthogonal ones, and the binding scale is the count of directions the output cloud genuinely occupies. With this refinement in place, the section turns to how the same compounding mechanism realises differently in each method.

2.3.4 DEA: Weight Optimisation

The DEA realisation is algorithmic, manifesting as weight optimisation. The linear program underlying DEA explicitly chooses each firm's weights to maximise its own efficiency score subject to the constraint that the efficiency score of the other $n-1$ peers exceeds unity under those weights (Charnes et al., 1978). A boundary firm with few binding constraints – few neighbors in the input-output space – can effectively ignore any dimension of the technology by assigning zero weight on a variable it performs poorly on, producing an efficient score that may rest on a single favourable, unique input-output combination relative the sample (Dyson and Thanassoulis, 1988). One important implication follows: whenever a variable is added, the LP can always reproduce a firm's previous efficiency score by assigning that variable zero weight, so efficiency scores can only weakly increase (Dyson et al., 2001). As the number of inputs and outputs grows with n fixed, the feasibility set and likelihood of a firm finding a favorable weight optimization increases (Dyson et al., 2001; Dyson and Thanassoulis, 1988). A common *rule-of-thumb* applied in practice is therefore that the sample needs at least $2p \times q$ firms to preserve discrimination (Dyson et al., 2001). As this benchmark is violated, the likelihood of firms becoming exterior self-evaluators (Edvardsen et al., 2008; Torgersen et al., 1996) increases and the methods ability to distinguish genuine efficiency decreases (Cook and Seiford, 2009). However, there is literature arguing for viewing the curse of dimensionality in intrinsic rather than nominal terms (see Adler and Yazhemsky, 2010; Jenkins and Anderson, 2003; Wilson, 2018)

The DEA pathology is more mature because the mechanism is directly visible in the dual constraints of the optimization problem and the literature has discussed it for decades (Cook and Seiford, 2009). However, the same can not be said about the corresponding mechanism in CNLS.

2.3.5 CNLS: Subgradient Overfitting

The CNLS realisation is statistical. The object is least squares subject to Afriat and monotonicity constraints, and for a boundary firm with few disciplinary neighbors the Afriat constraint admits slack and the LS objective uses that slack to drive the residual toward zero by selecting extreme slopes.

Asymptotically the CNLS estimator works on compact interior subsets where disciplinary neigh-

bors are plenty. Seijo and Sen (2011) prove that, in the multivariate case, the CNLS fitted cost level converges uniformly to the true cost function, and where the true function is differentiable the estimated slopes converge to the true marginal costs. Lim and Glynn (2012) confirm the consistency result and extend the result to the monotone convex case relevant for cost frontiers, noting however that uniform convergence on $\mathbb{R}^d$ typically fails because boundary subgradients may misbehave.

At the boundary the subgradient is unbounded in probability: the estimated marginal cost can be arbitrarily large in a given sample and does not converge to its true value as n increases. In the univariate case, Ghosal and Sen (2017) prove that the CNLS estimator is inconsistent at the boundary of its domain which Liao et al. (2026, Theorem 2) extend to the multivariate case: the minimum-norm subgradient is unbounded in probability, producing extreme slopes for firms near the corners of the input-output space and, more broadly, for those disciplined by few neighbors. Monotonicity alone does not remedy the problem, with Liao et al. (2026, Remark 4) noting that monotone CNLS is a degenerate case of weight-restricted convex regression with zero lower bound and infinite upper bound, allowing extreme slopes whenever doing so reduces the residual sum of squares.

2.3.6 Remedies: Bounding Weights and Slopes

The shared geometry of the frontiers and their flexibility mechanisms lends itself to their remedies rhyming, with both literatures attempting the same intervention: bounding the slope or weight vectors from going wild for boundary firms.

With the DEA flexibility realisation being algorithmic, manifesting as weight optimization, the literature primarily impose hard restrictions. The main approaches are absolute weight bounds (Dyson and Thanassoulis, 1988), virtual weight bounds and assurance regions (Sarrico and Dyson, 2004; Wong and Beasley, 1990), cross-efficiency with peer-averaged weights (Doyle and Green, 1994; Sexton et al., 1986), and production trade-offs re-expressing bounds as constraints on marginal rates of substitution (Podinovski, 2004, 2005). The aforementioned families of ex-ante remedies essentially share a common idea: no firm should be judged entirely by its own choosing. The trade-off formulation is conceptually appealing: Podinovski (2004) reframes weight restrictions as technology-space constraints on marginal rates of substitution, preserving the radial-improvement interpretation of efficiency that value-based weight restrictions often violate (Allen et al., 1997). All aforementioned restore discrimination, in theory, at the cost of an additional parameter in the estimation, reducing transparency in regulatory settings.

The CNLS remedies operates on the slope vector with soft and hard bounds in attempts to restore consistency at the boundary. Soft penalisation attempts to restore consistency by introducing a regularisation term on the slope vectors in the least-squares objective (Bertsimas and Mundru, 2021) rather than bounding them directly. Lim (2021, Theorem 1) proves consistency on compact interior subsets under the standard tuning $\lambda_n \to 0$, but Liao et al. (2024, Theorem 3) shows that the approach still leaves the minimum-norm subgradient unbounded in probability at the

boundary: soft penalisation does not remedy extreme subgradients for boundary firms.

In contrast, hard bounds on individual subgradients restore consistency across the entire domain, including at the boundary. Mazumder et al. (2019) is the canonical work here, introducing Lipschitz convex regression which bounds $\|\boldsymbol{\beta}_i\| \leq L$ for all firms. Liao et al. (2026) extend and generalise this in two directions and prove strong uniform consistency over the entire domain: Augmented Lipschitz Convex Regression (ALCR, Theorem 3) bounds deviations from reference subgradient $\|\boldsymbol{\beta}_i - \boldsymbol{\beta}_0\| \leq L$, and Weight-Restricted Convex Regression (WRCR, Theorem 5) imposes coordinate-wise bounds $\boldsymbol{\ell} \leq \boldsymbol{\beta} \leq \mathbf{u}$, where $\boldsymbol{\ell}, \mathbf{u} \in \mathbb{R}^d$, on the slope vectors. The Finnish Energy Authority has implemented weight-restricted CNLS in its DSO regulation since 2024, excluding the top and bottom deciles of the shadow-price distribution by constraints (Energiavirasto, 2023a). On the same EMV data, Liao et al. (2026) report that WRCR reduces out-of-sample RMSE by factor $\sim$70$\times$ relative to monotone CNLS.

3 Data

3.1 Regulatory benchmarking data

The primary dataset is Ei's official benchmarking dataset for Swedish local distribution network operators, compiled for the fourth regulatory period (RP4, 2024–2027). The dataset is cross-sectional, each firm is represented as a single observation in the data. The data is constructed and compiled from annual reports spanning 2018–2021 and regulatory capital base projections for 2024 (Energimarknadsinspektionen, 2024, *Bilaga 9*).

Input variables measure resource use, while output variables capture both what a DSO delivers to customers and the extent of the network it operates. All output variables and controllable operating costs are computed as the arithmetic mean across the four-year reference period of 2018–2021, preventing any single year from disproportionately affecting values. Capital costs are instead a forward-looking estimate for the first year of the regulatory period (2024) computed from the reported regulatory capital base as of 15 January 2024. Both controllable operating costs and capital costs are in 2022 price levels.

The dataset contains five outputs, and the distinction between them is central to the analysis that follows. Four are *service outputs*, measuring what a DSO delivers to its customers: subscriptions (CU), the number of customer connection points; peak demand (MW), the maximum subscribed capacity and actual peak demand from the overlying network; and energy delivered, split into high (MWh_h) and low (MWh_ℓ) voltage to reflect the different cost structures of the two voltage levels. The fifth output, network stations (NS), is different in kind: it counts the number of secondary transformer stations and proxies the physical extent of a firm's network rather than a service rendered to customers. Energimarknadsinspektionen (2024) reports that NS and total network length correlate at $r = 0.99$, so only one of the two is retained; the near-perfect correlation also underscores that NS is best read as a measure of network extent.

The original dataset contains 148 Swedish DSOs. In our analysis, we exclude four of these, leaving a final sample of 144 firms. Two of these reflect consolidation of firms through mergers: Hofors Elverk AB (REL00075) has been merged into Gävle Energi AB (REL03076), while Viggafors (REL03050) has been merged into Näckåns Elnät AB (REL03077). In both cases, costs and output variables are consolidated into the acquiring entity. We also exclude Carlfors Bruk (REL00024) and Sörbylunds Elnät (REL00965) due to them being micro firms with unique input-output combinations. These two have also been deemed as non-comparable by Energimarknadsinspektionen (2024). However, two other firms deemed as outliers by Ei: Övik Energi (REL00257) and Näckåns (REL03077), have been added back in due to StoNED being stochastic and less sensitive to outliers.

Table 3.1 reports summary statistics for the regulatory benchmarking variables across the 144

firms. The distribution exhibits substantial scale variation, reflecting the coexistence of small rural cooperatives and large national operators within the Swedish DSO landscape. The extreme right-skewness motivates the use of a multiplicative (log-linear) cost function specification.

Table 3.1: Summary statistics for the 144 Swedish DSOs in the analysis sample. Panel A reports cost and output variables: costs are in 2022 prices, with controllable operating costs averaged over 2018–2021 and capital costs as the forward-looking estimate for 2024. Panel B reports the ownership distribution. Source: Energimarknadsinspektionen (2024).

Panel A: Cost and output variables

Variable	Unit	Mean	SD	Min	p25	Median	p75	Max
TOTEX	MSEK	238.4	842.8	2.4	35.7	78.9	142.7	6 549.9
$OPEX_p$	MSEK	74.0	233.1	0.9	14.1	28.5	51.5	1 717.6
CAPEX	MSEK	164.3	610.4	1.5	21.9	51.6	93.4	4 832.3
CU	count	38 666	139 075	236	4 596	11 657	23 023	1 035 833
MW	MW	163.8	587.3	1.0	21.1	52.2	92.2	4 870.7
NS	count	1 256	5 539	22	154	296	579	46 039
MWh_ℓ	MWh	472 733	1 686 211	2 552	62 646	154 253	285 137	13 409 450
MWh_h	MWh	171 079	518 417	0	11 894	50 056	116 976	3 905 286

Panel B: Ownership

Category	N	Share (%)
Municipal	96	66.7
Private	22	15.3
Cooperative (ek. för.)	22	15.3
State	4	2.8

3.2 Output cloud geometry

The five output variables reported in Section 3.1 do not vary independently: almost all of their joint variation traces along a scale axis, with only NS consistently varying along a different axis. This section establishes the geometry and follows the dimensional discussion in Section 2.3.

Some housekeeping notes come first. Nine of the 144 firms report exactly zero high-voltage output, $MWh_h = 0$. Operating high-voltage transmission is a categorical operational distinction rather than a continuous output dimension, with the threshold an exact zero rather than a tuning parameter. We therefore restrict the geometric analysis in this section to the high-voltage operator sub-sample of $n = 135$; frontier estimation in Chapter 4 continues to use all 144 firms. Setting aside the nine zero-firms costs almost no information. A probit of the high-voltage indicator $\mathbf{1}\{MWh_h > 0\}$ on the four remaining log-outputs $\{\log CU, \log MW, \log NS, \log MWh_\ell\}$ classifies the pooled 144-firm sample with 95.1% accuracy (Appendix Table A.2). The zero-firms sit where the rest of the cloud predicts they should, and they carry no orthogonal information of their own.

Among the high-voltage operators, four of the five outputs move almost as one. The first principal component (PC1) of the standardised logs $\{\log CU, \log MW, \log MWh_\ell, \log MWh_h\}$ explains 92.0% of their joint variance: the four are nearly collinear and define a single dominant scale

axis. NS is the exception. Regressing $\log \mathrm{NS}$ on those same four outputs leaves $1 - R^2 = 0.205$ and tells us that a fifth of its variation varies along a different axis not explained by scale.

The effective rank r_e of the cloud tells the same story. For the covariance matrix Σ of the standardised five-output cloud, its effective rank r_e (Koltchinskii and Lounici, 2017; Vershynin, 2026) equals 1.119, barely above one. Dropping NS reduces it by only 0.032, while the same exercise on dropping CU, MW, or MWh_ℓ in turn instead *increases* effective rank (Appendix Table A.4. Of the five outputs, NS alone contributes a direction that the scale axis does not already contain.

A note may be in place for an observant reader of Appendix Table A.1 who notices that the MWh_h row of the full $1 - R^2$ grid reads 0.254 which exceeds that of NS's 0.205 and could therefore constitute a second orthogonal direction. However, the resemblence is superficial and we have to unfortunately disappoint our observent reader. The orthogonal variation in MWh_h is concentrated to a few firms rather than spread across the sample, and it behaves differently from NS's variation at the frontier. Separating geometric orthogonality from frontier relevance requires the frontier estimates themselves, so we defer the MWh_h puzzle to Chapter 6.

This pattern is the empirical statement of the refinement in Section 2.3.3: in finite samples the binding scale for boundary-share growth and escape directions is the effective rank r_e of the output cloud, not the nominal output count d. With $r_e \approx 1.12$ on five standardised log-outputs, the Swedish DSO output cloud occupies roughly one scale axis plus one orthogonal addition along NS.

3.3 Concession geography and fragmentation

Each system operator holds an exclusive network concession granting it the sole right to distribute electricity within a defined geographic area; the geographic boundaries of these are defined in Ei's official shapefile, which contains 311 polygons mapped to 144 firms, projected in EPSG:3006. Of the 144 firms, 126 are single-polygon firms while the remaining 18 are multi-polygon firms, firms whose concession areas are recorded as multiple polygons in the shapefile. Three large network operators – referred to as the *Big Three* – account for 46.6% of all polygons: E.ON with 70 polygons (22.5%), Vattenfall with 48 polygons (15.4%), and Ellevio with 27 polygons (8.7%). However, not all polygons belonging to the same firm are geographically separate. After dissolving adjacent polygons for the same firm, 124 of 144 firms (86%) have a single continuous territory, while the remaining 20 firms (14%) have genuinely disjoint concession areas. The *Big Three* are the most fragmented, with operating areas scattered across Sweden: E.ON operates in 10 disjoint parts, Vattenfall in 14, and Ellevio in 8. Table 3.2 summarises the concession-area shape and dispersion while Figure A.2 (Appendix A.2) maps the data to the Big Three.

Table 3.2: Descriptive statistics for concession-area geometry and centroid sensitivity.

Panel A: Concession geometry (N=144 firms, 311 polygon rows)							
Variable	Unit	N	Mean	SD	Min	Median	Max
Polygons (raw)	count	144	—	—	1	1	70
Disjoint components	count	144	—	—	1	1	14
Area	km^2	144	3 249	17 196	8	256	179 495
Compactness	—	144	0.413	0.154	0.022	0.413	0.826
Convexity	—	144	0.764	0.145	0.106	0.808	0.986
Bounding-box diagonal	km	144	71.1	165.5	4.3	33.1	1 471.9
Largest-component area share	—	20	0.766	0.150	0.442	0.801	1.000
Inter-component gap	km	20	5.9	5.7	0.4	5.4	23.5
Panel B: Centroid sensitivity (N=20 multi-component firms)							
Centroid shift, largest-area	km	20	91.5	198.8	1.2	10.4	684.0
Centroid shift, largest-pop	km	20	28.2	59.2	1.2	10.0	263.6
Centroid shift, polygon union	km	20	57.9	127.4	1.3	10.2	540.4
Max shift / bbox diagonal	—	20	0.247	0.167	0.050	0.156	0.603

3.4 Environmental z-variables

Controlling for structural heterogeneity in costs requires firm-level scalars of environmental conditions that affect distribution costs but lie outside firm control. We organise this into three dimensions, each tied to a distinct cost mechanism: *spatial dispersion*, reflecting how concentrated or scattered demand and infrastructure is and is currently proxied by Ei via NS; *vegetation*, reflecting forest-related costs such as tree fall and corridor clearing; and *climate*, reflecting seasonal operating costs from snow load. We proxy these dimensions through three z-variables – population density, forest cover, and snow exposure – at the firm level.

For each z-variable, the raw data are raster or grid covering Sweden and the question is how to aggregate that data into a single scalar per DSO. For compact single-polygon concessions, any reasonable aggregation rule yields essentially the same result. However, for multi-polygon firms with dispersed concessions, the aggregation rule is an explicit modelling choice with non-trivial consequences. This is the MAUP (see Section 2.2.2), and it propagates downstream to every object the z-vector touches. Section 3.4.1 develops the ex-ante principle that disciplines all three constructions.

3.4.1 Exposure-aligned weighting principle

Let cell i of a geographic raster carry a local value z_i of the structural factor under study, and let w_i be the exposure weight for the cost mechanism the z-variable is trying to capture. If cost contribution from cell i scales as $c_i \propto w_i \cdot z_i$, then the firm-level scalar consistent with that scaling is the w-weighted mean

$$\bar{z}_f = \frac{\sum_{i \in f} w_i \cdot z_i}{\sum_{i \in f} w_i}. \tag{3.1}$$

The proposed principle is motivated by: (i) it is an ex-ante restriction, the weighting and aggregation is derived from the theoretical cost mechanism and not determined ex-post from what

gives the desired outcome; and (ii) it allows and motivates heterogeneous weights for different z-variables.

The principle's determinacy varies across the z-variables. For population density, the relevant dimension is the density felt by the average resident, this leaves (Craig, 1984; Ottensmann, 2018) as the only operationally motivated aggregation. For snow and forest, the ideal weight would be line kilometers per cell, since the relevant cost mechanisms are load on conductors and vegetation-management cost per km of line. Line per km cell is however, to our knowledge, unobservable given publicly available data. Two reasonable proxy weights are available: population and area. Area-weighting assumes lines are uniformly distributed across concessions, which may overstate exposure in large, uninhabited rural concessions where the true exposure is lower. Population-weighting, on the other hand, assumes line dispersion scales with customers, which may be more reasonable for low-voltage networks, but may also understate exposure along high-voltage line infrastructure crossing sparse terrain between populated clusters. Neither proxy is unambiguously correct. Against this backbone, population-weighting serves as the main specification, preserving a single weighting principle across the z-vector. The area-weighted construction enters as MAUP-sensitive robustness for multi-polygon firms in Section 4.5.2.

3.4.2 Population density

For population density, population data is sourced from Statistics Sweden (SCB) 1 km×1 km population grid, vintage 2024-12-31, covering all populated areas of Sweden (115,062 cells) with zero-population cells being absent from the dataset. For each cell i that falls within the union of a firm's concession polygons, let pop_i be the cell's total population and $d_i = \text{pop}_i/1\,\text{km}^2$ its local density. The main specification is then population-weighted density (PWD) (Craig, 1984; Ottensmann, 2018) computed as

$$d^{\text{PWD}}(P) \;=\; \frac{\sum_{i\in P} \text{pop}_i\, d_i}{\sum_{i\in P} \text{pop}_i}, \tag{3.2}$$

measuring the density experienced by the average customer, and is the operationally relevant quantity: a concession dominated by uninhabited forest can have near-zero arithmetic density

$$d^{\text{arith}}(P) \;=\; \frac{\sum_{i\in P} \text{pop}_i}{\sum_{i\in P} a_i} \tag{3.3}$$

despite serving customers in dense clusters within smaller areas of the concession. PWD – and population-weighted forest and snow exposure – is computed over the union of the firm's grid cells, ignoring which polygon each cell belongs to,

$$d_f^{\text{PWD}} \;=\; d^{\text{PWD}}\left(\bigcup_p P_{f,p}\right). \tag{3.4}$$

The alternative, and more intuitive, approach is to compute $d^{\text{PWD}}(P_{f,p})$ for each polygon p and then combine across polygons via population weighting. However, for a firm f whose polygons $\{P_{f,p}\}_{p=1}^{k_f}$ partition a set of grid cells, the firm-level PWD computed over the union of cells equals

the population-weighted aggregation of polygon-level PWDs.

An alternative aggregation, used as a MAUP comparator for the Big Three in Section 4.5.2, replaces the polygon-population weights with polygon-area weights:

$$d_f^{\text{poly-area}} \;=\; \frac{\sum_p a_{f,p}\, d^{\text{PWD}}(P_{f,p})}{\sum_p a_{f,p}}. \tag{3.5}$$

Unlike the population-weighted polygon aggregation, this generally differs from d_f^{PWD}, since polygons with the same area can carry very different populations.

3.4.3 Snow

Snow data is sourced from the Swedish Meteorological and Hydrological Institute (SMHI) on a 2.5 km grid in Lambert Conformal Conic projection, spanning 336 monthly observations from January 1991 to December 2018. Grid cells are reprojected to EPSG:3006 and spatially joined to concession polygons. The main snow variable is the mean annual maximal depth

$$s_i \;=\; \frac{1}{Y}\sum_{y=1}^{Y} \max_{t\in \text{SY}_y} \text{snd}_{i,t}, \tag{3.6}$$

with firm-level aggregation

$$\bar{s}_f^{\text{pop}} \;=\; \frac{\sum_{i\in f} \text{pop}_i\, s_i}{\sum_{i\in f} \text{pop}_i}, \tag{3.7}$$

to capture the expected peak load on overhead infrastructure under population-weighted specifications.

As a complementary statistic used in the MAUP sensitivity panel for the Big Three (Section 4.5.2), we also compute the mean annual count of snowdays above threshold $\tau = 0.3$

$$s_i^{\text{days}} \;=\; \frac{1}{Y}\sum_{y=1}^{Y} \left|\{t \in \text{SY}_y : \text{snd}_{i,t} > \tau\}\right|, \tag{3.8}$$

Firm-level aggregation follows the population-weighted form above, with s_i replaced by s_i^{days}.

3.4.4 Forest

Forest data is sourced from the 2018 NMD raster produced by Lantmäteriet at 10 m×10 m resolution and 2018 vintage. We apply binary forest classification based on raster classifications in the data: 16 NMD codes corresponding to forest (pine, spruce, mixed etc.) are classified as forest, while all other codes (open wetland, arable, water) are classified as non-forest. Temporarily non-forested areas are classified as forest, since infrastructure in these areas was likely planned and built under forested conditions which are reflected in costs. Population-weighted forest cover

is then computed as the percentage of each polygon classified as forest:

$$\bar{F}_f^{\text{pop}} \;=\; \frac{\sum_{i \in f} \text{pop}_i \, F_i}{\sum_{i \in f} \text{pop}_i}. \tag{3.9}$$

4 Methodology

4.1 Specification context

Regulatory benchmarking requires a decision about where structural cost drivers enter the model. Ei's DEA specification places NS in the output vector alongside the four service outputs (CU, MW, MWh_ℓ, MWh_h) and each firm receives a set of weights to maximise its efficiency score subject to constraints. While the four service outputs measure what the firm delivers to its customers, NS is intended to credit firms for operating in sparse areas. Whether that credit actually reaches the firms whose density warrants it is the subject of RQ1. StoNED provides an alternative way of handling structural cost drivers: structural variables can enter through a separate parametric channel that adjusts each firm's cost level without changing the shape of the frontier. This cleanly separates the production question (how efficiently does the firm use its resources?) from the structural question (how costly is the environment it operates in?). Whether this alternative channel changes the efficiency scores of firms where NS credit doesn't align with the firm's operating environment is the subject of RQ2.

Ei currently applies OPEX and CAPEX as separate inputs with five outputs. For the upcoming regulatory period (RP5), Ei will switch to TOTEX, merging the two cost elements into a single aggregate. The main specifications in this thesis therefore use TOTEX as the cost input, consistent with the RP5 transition. Table 4.1 reports the full set of specifications for both research questions.

4.2 Data Envelopment Analysis (DEA)

We use the input-based DEA model with constant returns to scale (CRS), which is the CCR model after the names of its developers Charnes, Cooper, and Rhodes (1978). This is Ei's specification and the one we base our empirical analysis on; we retain it deliberately, despite the objection that CRS bundles scale with managerial efficiency.[1]

DEA uses the following linear programme to determine the efficiency of a particular firm:

$$\theta_0^* = \min \theta \tag{4.1}$$

[1] An efficiency score under CRS is often interpreted as "overall technical efficiency", which decomposes multiplicatively into pure technical (managerial) efficiency and scale efficiency. Under CRS, a firm operating with managerial best practice but away from its most productive scale still scores below one. VRS removes this scale-slack by adding a convexity constraint ($\sum_j \lambda_j = 1$) that benchmarks each firm only against others of comparable size. We retain the scale-slack on the basis that this is a lever through which Ei incentivises consolidation through mergers (Energimarknadsinspektionen, 2024), consistent with the unexploited economies of scale for the Swedish sector observed in the literature (for a general treatment, see Söderberg and Vesterberg, 2025)

subject to

$$\sum_{j=1}^{n} \lambda_j x_{ij} \leq \theta x_{io}, \quad i = 1, \ldots, m, \tag{4.2}$$

$$\sum_{j=1}^{n} \lambda_j y_{kj} \geq y_{ko}, \quad k = 1, \ldots, s, \tag{4.3}$$

$$\lambda_j \geq 0, \quad j = 1, \ldots, n, \tag{4.4}$$

where x_{ij} is input i consumed by firm j and y_{kj} is output k of firm j. The first constraint requires the convex combination of peers to consume no more of each input than the fraction θ of the firm's observed input bundle. The second requires the peer combination to produce at least as much of each output. The third restricts the intensity weights λ_j to be non-negative, so the reference point is fully built from observed firms.

The efficiency score $\theta_0^* \in [0, 1]$ measures how much firm 0 could reduce its cost while still producing the same output, given the performance of other firms. The weight vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$ contains the intensity weights that determine which firms are combined to create firm 0's reference set.

DEA has a second formulation, the multiplier problem, in which each firm selects output weights $\mu_k \geq 0$ and an input weight $v \geq 0$ to maximise its own efficiency ratio, subject to the constraint that no firm achieves a ratio above one under those same weights. Solving the two jointly yields both the intensity weights $\boldsymbol{\lambda}$, which identify the peer reference set, and the output weights μ_k, which determine how much credit each output receives. From the latter, virtual outputs $\mu_k y_{ik}$ and virtual output shares $s_{ik} = \mu_k y_{ik} / \sum_j \mu_j y_{ij}$ are derived.

Since $\mu_k \geq 0$ for every output, a firm can always set $\mu_{\text{NS}} = 0$ and be evaluated without reference to NS. Whether NS actually enters a firm's efficiency rating depends on whether the optimisation finds it beneficial to do so, which differs across firms. NS as structural compensation therefore does not mechanically reach every firm.

4.3 Stochastic nonparametric envelopment of data (StoNED)

4.3.1 StoNED

StoNED estimates a nonparametric, deterministic frontier from the data, then decomposes the residuals into a non-negative inefficiency term and a symmetric noise term. This thesis uses the multiplicative form

$$\ln C_i = \ln C(\mathbf{y}_i) + u_i + v_i, \tag{4.5}$$

where $u_i \geq 0$ (half-normal) is the inefficiency term and $v_i \sim N(0, \sigma_v^2)$ is random noise; this composed-error structure follows the stochastic frontier specification of Aigner et al. (1977) and Meeusen and Den Broeck (1977). The multiplicative specification is the standard choice in

applied StoNED work and ensures consistency with CRS. With additive errors, the best-practice frontier is obtained by shifting the estimated average-practice frontier downward by expected inefficiency $E[u_i] = \sigma_u\sqrt{2/\pi}$, but this shift can violate constant returns to scale (Kuosmanen and Kortelainen, 2012). Besides the distributional assumptions, independence between u_i, v_i, and outputs is also assumed.

Step 1, CNLS. The frontier is estimated by solving a constrained least-squares problem. In log-linear form,

$$\min_{\hat{\varphi},\beta} \sum_{i=1}^{n} (\ln C_i - \ln \hat{\varphi}_i)^2 \tag{4.6}$$

subject to

$$\hat{\varphi}_i = \boldsymbol{\beta}_i' \mathbf{y}_i, \quad \forall i = 1, \ldots, n, \tag{4.7}$$

$$\boldsymbol{\beta}_i' \mathbf{y}_i \geq \boldsymbol{\beta}_h' \mathbf{y}_i, \quad \forall h, i = 1, \ldots, n, \tag{4.8}$$

$$\boldsymbol{\beta}_i \geq 0, \quad \forall i = 1, \ldots, n. \tag{4.9}$$

Each firm receives its own slope vector $\boldsymbol{\beta}_i = (\beta_{i,1}, \ldots, \beta_{i,q})$, the firm-specific vector of marginal costs. The second constraint (Afriat inequalities) enforces global convexity by requiring that each firm's fitted hyperplane lies weakly above every other firm's hyperplane when evaluated at the same output point. The third constraint enforces monotonicity, making sure that higher output never lowers costs. The resulting frontier is piecewise-linear, with each firm potentially defining its own segment. Under the multiplicative CRS specification used throughout the thesis, the intercepts $\alpha_i = 0$ for all i, so the hyperplane passes through the origin and is fully determined by the slope vector: $\hat{\varphi}_i = \boldsymbol{\beta}_i' \mathbf{y}_i$. Geometry and coefficients therefore coincide. CNLS residuals $\hat{\varepsilon}_i = \ln C_i - \ln \hat{\varphi}_i$ capture the composite error $u_i + v_i$ but do not yet separate inefficiency from noise.

Step 2, Decomposition. The variance parameters σ_u (inefficiency) and σ_v (noise) are recovered from the CNLS residuals $\hat{\varepsilon}_i$ using quasi-likelihood estimation (Fan et al., 1996, QLE). QLE maximises a pseudo log-likelihood over the signal-to-noise ratio $\kappa = \sigma_u/\sigma_v$

$$\hat{\kappa} = \arg\max_{\kappa} \sum_{i=1}^{n} \ln\left[\frac{2}{\sigma}\,\phi\left(\frac{\hat{\varepsilon}_i}{\sigma}\right)\Phi\left(\frac{\kappa\,\hat{\varepsilon}_i}{\sigma}\right)\right], \qquad \sigma^2 = \sigma_u^2 + \sigma_v^2, \tag{4.10}$$

where $\hat{\sigma}_u$ and $\hat{\sigma}_v$ are recovered from $\hat{\kappa}$ and the residual variance. Separation is possible because inefficiency is one-sided so if the residuals are sufficiently skewed, the data contain enough information to distinguish the two components. Firm-level efficiency scores are recovered using the Battese and Coelli (1988) conditional expectation

$$\widehat{\mathrm{TE}}_i = \frac{\Phi(\mu_i^*/\sigma^* - \sigma^*)}{\Phi(\mu_i^*/\sigma^*)} \exp\!\left(-\mu_i^* + \tfrac{1}{2}(\sigma^*)^2\right), \tag{4.11}$$

with $\sigma^* = \sigma_u\sigma_v/\sqrt{\sigma_u^2 + \sigma_v^2}$ and $\mu_i^* = \sigma_u^2/(\sigma_u^2 + \sigma_v^2) \cdot \hat{\varepsilon}_i$ in the cost orientation. By construction $\widehat{\mathrm{TE}}_i \in (0, 1]$, the same scale as DEA's θ_i, so the two methods' scores are directly comparable.

The Battese and Coelli (1988) predictor of technical efficiency is preferred to exponentiating the Jondrow et al. (1982) inefficiency predictor $\exp(-E[u_i \mid \hat{\varepsilon}_i])$ because $\exp(\cdot)$ is convex: by Jensen's inequality $\exp(-E[u_i \mid \hat{\varepsilon}_i]) \neq E[\exp(-u_i) \mid \hat{\varepsilon}_i]$, and only the latter is the minimum mean-squared-error predictor of the efficiency $\exp(-u_i)$ we are trying to estimate.

4.3.2 StoNEZD

Environmental variables enter the model through a one-stage parametric extension developed by Johnson and Kuosmanen (2011). The cost frontier becomes:

$$\ln C_i = \ln C(\mathbf{y}_i) + \boldsymbol{\delta}'\mathbf{z}_i + u_i + v_i, \tag{4.12}$$

where $u_i \geq 0$ is the inefficiency term. On levels, the model becomes

$$C_i = C(\mathbf{y}_i) \cdot \exp(\boldsymbol{\delta}'\mathbf{z}_i) \cdot \exp(u_i + v_i), \tag{4.13}$$

confirming that the multiplicative label refers to the error structure, not the z-entry. The CNLS objective is modified to:

$$\min_{\boldsymbol{\beta},\boldsymbol{\delta},\varphi} \sum_{i=1}^{n} (\ln C_i - \ln \varphi_i - \boldsymbol{\delta}'\mathbf{z}_i)^2 \tag{4.14}$$

subject to

$$\varphi_i = \boldsymbol{\beta}_i'\mathbf{y}_i, \quad \forall i = 1, \ldots, n, \tag{4.15}$$

$$\boldsymbol{\beta}_i'\mathbf{y}_i \geq \boldsymbol{\beta}_h'\mathbf{y}_i, \quad \forall h, i = 1, \ldots, n, \tag{4.16}$$

$$\boldsymbol{\beta}_i \geq 0, \quad \forall i = 1, \ldots, n, \tag{4.17}$$

with the same convexity and monotonicity constraints as before. The coefficient vector $\boldsymbol{\delta}$ is common to all firms, meaning environmental variables shift the frontier up or down without changing its shape. Johnson and Kuosmanen (2011) shows that $\hat{\boldsymbol{\delta}}$ is consistent, asymptotically efficient, and converges at the standard parametric rate $n^{-1/2}$, so conventional statistical inference applies to the z-variables even though the frontier itself is estimated nonparametrically. Computationally, $\hat{\boldsymbol{\delta}}$ is equivalent to regressing the CNLS log-cost residuals on $\mathbf{z}$ by OLS (Kuosmanen and Johnson, 2010); see equation (4.27) in Section 4.5.2. Crucially, $\mathbf{z}_i$ does not enter the nonparametric component, so adding environmental variables does not increase the dimensionality of the frontier and does not trigger the curse of dimensionality that affects DEA.

4.4 Estimation design

4.4.1 Specifications

The full set of specifications is reported in Table 4.1. Panel A specifications are estimated for both DEA and StoNED. Panel B specifications (StoNEZD) are estimated at both $k = 4$ and $k = 5$ to separate the effect of introducing the z-channel from the effect of removing NS from the output vector.

Table 4.1: Specifications for RQ1 and RQ2

k	Input	Outputs	z-variables
Panel A: RQ1 specifications (DEA and StoNED)			
5	TOTEX	CU, MW, NS, MWh_ℓ, MWh_h	—
4	TOTEX	MW, NS, MWh_ℓ, MWh_h	—
4	TOTEX	CU, NS, MWh_ℓ, MWh_h	—
4	TOTEX	CU, MW, MWh_ℓ, MWh_h	—
4	TOTEX	CU, MW, NS, MWh_h	—
4	TOTEX	CU, MW, NS, MWh_ℓ	—
Panel B: RQ2 specifications (StoNEZD)			
4	TOTEX	CU, MW, MWh_ℓ, MWh_h	PWD
4	TOTEX	CU, MW, MWh_ℓ, MWh_h	PWD, Snow depth (pop-wt.), Forest cover (pop-wt.)
4	TOTEX	CU, MW, MWh_ℓ, MWh_h	PWD, Snow depth (area-wt.), Forest cover (area-wt.)
4	TOTEX	CU, MW, MWh_ℓ, MWh_h	PWD, Snow days (pop-wt.), Forest cover (pop-wt.)
5	TOTEX	CU, MW, NS, MWh_ℓ, MWh_h	PWD
5	TOTEX	CU, MW, NS, MWh_ℓ, MWh_h	PWD, Snow depth (pop-wt.), Forest cover (pop-wt.)
5	TOTEX	CU, MW, NS, MWh_ℓ, MWh_h	PWD, Snow depth (area-wt.), Forest cover (area-wt.)

4.5 Diagnostics

4.5.1 Diagnosing NS as a compensation instrument (RQ1)

Output-weight distribution. For each method, the per-firm component $\mu_{ik}y_{ik}$ (DEA virtual outputs, (Sarrico and Dyson, 2004; Thanassoulis et al., 2008)) and $\beta_{ik}y_{ik}$ (StoNED contributions in raw cost units) are computed. This is not to be mistaken for the weights the LP solver (shadow prices) puts on each variable for each firm. Row-normalising yields the firm-level virtual-share matrix (DEA) and contribution-share matrix (StoNED)

$$s_{ik}^{\text{DEA}} = \frac{\text{VO}_{ik}}{\sum_j \text{VO}_{ij}}, \qquad s_{ik}^{\text{StoNED}} = \frac{C_{ik}}{\sum_j C_{ij}}, \tag{4.18}$$

which sum to one across outputs and measure each variable's relative contribution in the frontier estimation. The raw multipliers μ (DEA shadow prices) and slopes β (CNLS marginal costs) are not comparable *in levels* because they are normalized differently. The share matrix s_{ik} is however unit-free and sum to one across outputs, making the *relative importance* of each output in the frontier estimation comparable across methods. Four readings are reported per output, side by side for DEA and StoNED.

Mean share, $\bar{s}_k = n^{-1}\sum_i s_{ik}$ measures the average reliance of a firm on output k.

Outside-whisker (%), the percentage of the $n = 144$ firms whose virtual / contribution share lies

outside $[Q_1 - 1.5\,\mathrm{IQR},\ Q_3 + 1.5\,\mathrm{IQR}]$. A heavy outlier mass for a given variable may indicate that the piecewise-linear frontier is bending sharply in that dimension to accommodate local observations.

Zero-weight (%), the percentage of firms placing a near-zero weight on that specific output. A zero DEA multiplier means that the linear program has found it optimal to ignore output k for firm i, while a zero StoNED weight means that the CNLS solution pushed firm i's marginal cost on dimension k to the monotonicity constraints limit. The mechanism differs, but the consequence is similar: a certain output k plays no role in determining the firm's efficiency score.

Variance decomposition, the variance share

$$\mathrm{share}_k^B = \frac{\mathrm{Var}_i(\mathrm{comp}_{ik})}{\sum_j \mathrm{Var}_i(\mathrm{comp}_{ij})} \tag{4.19}$$

sums to one across outputs and reads as the cross-firm variation in fitted cost attributable to variable k. The substitution ratio:

$$\mathrm{ratio}_k^A = \frac{\mathrm{Var}_i(\mathrm{comp}_{ik})}{\mathrm{Var}_i\left(\sum_j \mathrm{comp}_{ij}\right)} \tag{4.20}$$

reads as the share of each component's variance relative to the variance of the row sum. It sums to one when outputs vary independently, exceeds one under negative covariance (substitution: when one component drives the row sum, others fade), and falls below one under positive covariance (complementarity).

Leave-one-variable-out. To assess each output's contribution to the estimated frontier, we apply a LOVO (leave-one-variable-out) exercise (Lei et al., 2018) to both DEA and StoNED. For each single-output drop $k \in \{\mathrm{CU}, \mathrm{MW}, \mathrm{NS}, \mathrm{MWh}_\ell, \mathrm{MWh}_h\}$, we report the following diagnostics relative to the $k = 5$ specification. For DEA: the number of fully efficient firms, mean efficiency $\bar{\theta}$, the largest absolute change in θ_i, and the Spearman rank correlation. For StoNED: mean Battese-Coelli efficiency $\overline{\text{BC-TE}}$, Spearman rank correlation, and the mean and maximum absolute log-frontier shift:

$$\Delta_i^{(-k)} = \ln \hat{\varphi}_i^{(k=5)} - \ln \hat{\varphi}_i^{(-k)}, \tag{4.21}$$

where

$$\hat{\varphi}_i = \boldsymbol{\beta}_i' \mathbf{y}_i \tag{4.22}$$

defines the deterministic average-practice cost frontier against which firms are benchmarked.

NS–density alignment. For each firm, the NS share of its total efficiency score is computed: the virtual NS share $\mu_{\mathrm{NS}} y_{\mathrm{NS}} / \sum_k \mu_k y_k$ in DEA and the NS contribution share $\beta_{\mathrm{NS}} y_{\mathrm{NS}} / \sum_j \beta_j y_j$ in StoNED. The relationship between NS share and log population-weighted density is then

estimated via a fractional-response regression with logit link (Papke and Wooldridge, 1996)

$$E\left[s^{\bullet}_{i,\mathrm{NS}} \mid \log \mathrm{PWD}_i\right] = \Lambda(\alpha + \beta \log \mathrm{PWD}_i), \qquad \Lambda(t) = \frac{e^t}{1+e^t}, \tag{4.23}$$

separately for DEA and StoNED, fitted by Bernoulli quasi-MLE with robust standard errors. The fitted curve $\hat{p}_i = \Lambda(\hat{\alpha} + \hat{\beta} \log \mathrm{PWD}_i)$ traces expected NS share as a function of density across both methods. Firms that deviate from this expected relationship are identified and ranked by their standardised Pearson residual:

$$r_i^P = \frac{s^{\bullet}_{i,\mathrm{NS}} - \hat{p}_i}{\sqrt{\hat{p}_i(1-\hat{p}_i)}}. \tag{4.24}$$

The standardised Pearson residual measures how far each firm's observed NS share departs from the model's prediction, scaled by the variance implied by the logistic curve. Dividing by $\sqrt{\hat{p}_i(1-\hat{p}_i)}$ accounts for the heteroskedasticity inherent in fractional outcomes and makes residuals comparable across firms at different density levels.

Firms with $r_i^P \gg 0$ carry more NS weight than their density predicts (candidates for *wrong-firm compensation*); firms with $r_i^P \ll 0$ carry less than their density predicts (candidates for *missed-firm compensation*). The top five firms in each direction per method form the case-study set carried into the results chapter. Robustness is assessed by adding snow and forest cover and allowing a quadratic term in log density.

4.5.2 Evaluating the parametric z-channel (RQ2)

Under RQ2, the failure modes identified in RQ1 carry over to the StoNEZD specification with the four service outputs and a population-weighted density (PWD) z-variable. Structural heterogeneity is guided through a single global shift coefficient $\hat{\delta}$ (a scalar in the single-z headline; we write $\hat{\boldsymbol{\delta}}$ for the corresponding vector in full-z specifications, of which $\hat{\delta}$ is the first component) rather than through firm-specific NS slopes. Each failure mode implies a directional prediction for whether the z-channel relocates structural compensation in the right direction. Three diagnostics follow the log-cost decomposition below: $\hat{\delta}$ inference, cohort falsification, and MAUP sensitivity for the Big Three.

Log-cost decomposition. Under the main StoNEZD specification, fitted log-cost (the model's prediction of $\ln C_i$ for each firm) decomposes additively into an output component and an environmental component,

$$\ln \hat{C}_i = \ln \hat{\varphi}_i + \boldsymbol{\delta}' \mathbf{z}_i, \tag{4.25}$$

where $\hat{\varphi}_i = \boldsymbol{\beta}_i' \mathbf{y}_i$ is the deterministic average-practice frontier and $\boldsymbol{\delta}$ is the global shift vector. Equivalently, on levels,

$$\hat{C}_i = \hat{\varphi}_i \cdot \exp\left(\boldsymbol{\delta}' \mathbf{z}_i\right), \tag{4.26}$$

making the environmental component a multiplicative shift on the output-component frontier. The first two diagnostics below live in this additive log space.

$\hat{\delta}$ inference. Statistical inference is reported for the StoNEZD z-coefficients across both single-z and full-z specifications. The computation of $\hat{\boldsymbol{\delta}}$ in the CNLS quadratic program is numerically equivalent to regressing the log-cost residuals on the z-variables in a second-stage OLS,

$$\hat{\boldsymbol{\delta}}^{\,\mathrm{OLS}} = \big(\mathbf{Z}'\mathbf{Z}\big)^{-1}\mathbf{Z}'\big(\ln \mathbf{C} - \ln \hat{\boldsymbol{\varphi}}(\mathbf{Y})\big), \tag{4.27}$$

with robust standard errors, which account for potential heteroskedasticity in the CNLS residuals. We estimate $\hat{\boldsymbol{\delta}}$ across several density measures (linear PWD, log PWD, arithmetic-mean density, polygon-area-weighted density) in single-z specifications as well as for the full z-vector under population weighting, to test whether the PWD coefficient is robust to the inclusion of snow and forest covariates.

Cohort falsification. Cohort falsification tests wrong-firm and missed-firm compensation by computing two scalars per firm: a log-cost shift ΔC_i measuring the difference in fitted log-cost between the specifications, and an efficiency-score shift ΔT_i recording the change in the firm's efficiency score. An efficiency-score shift is only expected when ΔC_i is heterogeneous across firms: if the specification change shifts all firms' fitted costs by the same amount, relative ranks are unchanged and $\Delta T_i \approx 0$.

The output-component contribution that NS adds to fitted cost under the $k = 5$ StoNED specification is captured by the NS-LOVO log shift:

$$\Delta_i^{(-\mathrm{NS})} = \ln \hat{\varphi}_i^{(5)} - \ln \hat{\varphi}_i^{(4,\,-\mathrm{NS})}, \tag{4.28}$$

where $\hat{\varphi}_i^{(4,-\mathrm{NS})} = \sum_{j\neq \mathrm{NS}} \hat{\beta}_{ij}^{(4,-\mathrm{NS})} y_{ij}$ is the deterministic frontier from a separate $k = 4$ StoNED fit that drops NS from the output vector.

The environmental component in StoNEZD is the additive term $\hat{\boldsymbol{\delta}}'\mathbf{z}_i$, which collapses to $\hat{\delta}\,\mathrm{PWD}_i$ in the single-z headline ($k = 1$) and to $\sum_k \hat{\delta}_k\, z_{i,k}$ in the full-z anchor. The two contributions combine into

$$\Delta C_i = \hat{\boldsymbol{\delta}}'\mathbf{z}_i \;-\; \Delta_i^{(-\mathrm{NS})}, \tag{4.29}$$

which, in logs, reads as the firm-specific net fitted-cost difference between the two specifications. The difference in efficiency scores,

$$\Delta T_i = \widehat{\mathrm{TE}}_i^{\,\text{StoNEZD-4+}z} - \widehat{\mathrm{TE}}_i^{\,\text{StoNED-5}}, \tag{4.30}$$

is then recorded. The wrong-firm candidates (over-NS rows in Table 5.6), whose NS weights exceed what PWD predicts, should populate the south-west quadrant ($\Delta C_i < 0,\ \Delta T_i < 0$): routing density through the z-channel adds less to their fitted cost than NS previously did, so they are penalised relative to the NS-based specification. The missed-firm candidates (under-NS rows in Table 5.6), whose NS weights fall below what PWD predicts, should populate the north-east quadrant ($\Delta C_i > 0,\ \Delta T_i > 0$) for the opposite reason.

MAUP sensitivity. As a robustness check, we report MAUP sensitivity for the Big Three (Vattenfall, E.ON, Ellevio) across alternative aggregation choices, covering fitted cost and technical efficiency. The single-z panel varies the PWD specification across linear (reference), log, arithmetic-mean density, and polygon-area-weighted aggregation. The full-z panel varies the aggregation across population-weighted (reference), area-weighted, and an alternative snow-days measure (3.8).

5 Results

5.1 Network Stations as Implicit Structural Compensation (RQ1)

5.1.1 How DEA and StoNED distribute output weight

Table 5.1: Output-share, zero-weight, and variance-decomposition summary, k=5.

	Virtual/Contribution shares						Variance decomposition			
	Mean		> whisker (%)		Zero-weight (%)		Variance share		Substitution	
Output	DEA	StoNED	DEA	StoNED	DEA	StoNED	DEA	StoNED	DEA	StoNED
CU	0.328	0.158	0.0	6.2	14.6	0.0	0.379	0.231	2.53	0.08
MW	0.083	0.084	2.8	2.8	31.2	0.0	0.046	0.020	0.31	0.01
NS	0.230	0.273	0.0	3.5	4.9	0.0	0.162	0.438	1.09	0.15
MWh_ℓ	0.313	0.448	0.0	0.0	43.8	0.0	0.349	0.309	2.33	0.10
MWh_h	0.046	0.036	17.4	18.1	81.9	0.7	0.065	0.001	0.43	0.00
Sum	1.000	1.000	—	—	—	—	1.000	1.000	6.69	0.34

Table 5.1 reports how each method distributes weight across the five outputs. The most significant difference is not in the averages but in the extremes. DEA assigns zero weight to at least one output for the majority of firms: 82 percent of firms receive no credit at all for high-voltage deliveries, 44 percent receive none for low-voltage deliveries, and 31 percent receive none for subscribed power. In practice, this means the linear program has found it optimal to score these firms as if entire dimensions of their service do not exist. StoNED produces essentially no "zero weights" (one firm on MWh_h), because its shape constraints prevent the local slopes from collapsing in the same way. The above-whisker column (firms with outlier output shares) shows broadly similar rates across methods, with StoNED producing slightly more outliers on CU and NS. Shares are not comparable across methods (0.3 in DEA does not carry the same meaning as 0.3 in StoNED), but within each method they provide a consistent unit-free measure of relative output importance.

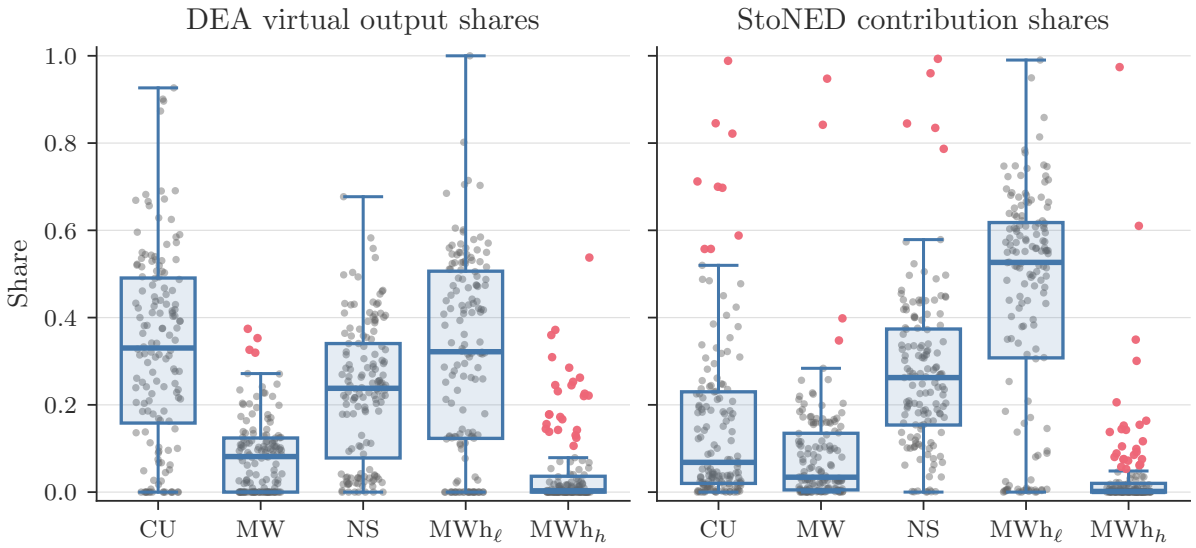


Figure 5.1: Distribution of virtual output shares (DEA) and contribution shares (StoNED), k=5 TOTEX baseline.

The box-plots in Figure 5.1 make the distributional contrast in shares visible. DEA shares span the full zero-to-one range for CU, with wider interquartile ranges reflecting the LP's freedom to concentrate weight. StoNED's interquartile ranges are narrower, but the method produces more individual outliers: 9 firms outside the range on CU and 5 on NS, compared to zero for DEA on both. The pattern suggests that DEA distributes its flexibility broadly across all firms through its unconstrained weight selection, while StoNED's shape constraints compress the majority but force a small number of firms into extreme local slopes to accommodate their observations.

5.1.2 Network stations as the dominant flexibility driver

Table 5.2: Leave-one-variable-out summary, k=5 baseline.

	DEA				StoNED			
Specification	n_{eff}	$\bar{\theta}$	ρ	max $\lvert\Delta\theta\rvert$	$\overline{\text{BC-TE}}$	ρ	mean $\lvert\Delta \ln \varphi\rvert$	max $\lvert\Delta \ln \varphi\rvert$
Baseline (k=5)	16	0.830	—	—	0.796	—	—	—
Drop CU	11	0.785	0.910	0.301	0.798	0.976	0.018	0.178
Drop MW	13	0.820	0.979	0.218	0.800	0.986	0.014	0.147
Drop NS	**6**	**0.735**	**0.758**	**0.472**	**0.758**	**0.643**	**0.126**	**0.779**
Drop MWh$_\ell$	10	0.799	0.941	0.267	0.785	0.916	0.045	0.309
Drop MWh$_h$	13	0.814	0.938	0.236	0.795	0.991	0.008	0.106

Table 5.2 reports what happens when each output is removed in turn from the k=5 baseline. In both methods, NS is the dominant driver of model flexibility by a wide margin. In DEA, removing NS cuts the number of efficient firms from 16 to 6, reduces mean efficiency by 9.5 percentage points, and produces a maximum individual score change of 0.472. No other single output comes close: the next most disruptive drop (CU) removes 5 efficient firms and shifts mean efficiency by 4.5 percentage points. In StoNED, the same pattern holds through different

diagnostics. Removing NS reduces mean BC TE by 3.6 percentage points, produces the lowest rank correlation against the baseline (0.643), the largest mean log-cost shift (0.126), and the largest individual cost shift (0.779). No other single drop comes close: all other rank correlations remain above 0.91.

5.1.3 NS dependence, rurality, and peer structure

Table 5.3: Retainers vs. droppers among DEA-efficient firms when NS is removed.

Panel A. Macro covariation with rurality (n = 144)			
	Pearson r	Spearman ρ	Expected sign
log_pwd × DEA NS virtual share	−0.73	−0.74	−
log_pwd × StoNED NS contribution share	−0.74	−0.71	−
log_pwd × $\Delta\theta_{\text{DEA}}$ on NS drop	+0.71	+0.68	+
log_pwd × ΔBC-TE$_{\text{StoNED}}$ on NS drop	+0.72	+0.64	+
Panel B. Within DEA-efficient frontier (n = 16; 10 of 16 lose efficient status when NS is dropped)			
	Retainers (n = 6)	Droppers (n = 10)	
Mean NS virtual share	0.05	0.32	
Retainers with non-NS dominant share ≥ 0.41	6/6	—	
Mean peer frequency	9.7	238.8	

Table 5.3 connects the LOVO (Leave-one-variable-out) results to firms' operating environments. The 16 DEA-efficient firms split cleanly when NS is removed; six retain their efficient status, and ten lose it. The two groups differ sharply; retainers carry an average NS share of 0.052 and anchor their frontier position on CU or MWh$_\ell$, while droppers carry an average NS share of 0.322 and tend to be more rural (mean log population-weighted density of 6.41 versus 7.67). Droppers are also far more frequently cited as benchmark peers, with a mean peer frequency of 238.8 compared to 9.7 for retainers. Removing NS, therefore, does not just change the scores of ten firms; it restructures the peer landscape for the entire sample, because the most widely used reference points in the sample are the exact firms whose efficiency status depends on NS.

The pattern extends beyond the efficient subset. Across all firms, rural operators carry more NS weight (Spearman $\rho = -0.735$ in DEA, -0.710 in StoNED) and suffer larger efficiency losses when NS is removed ($\rho = +0.684$ in DEA, $+0.643$ in StoNED). NS compensation is therefore not random; at the aggregate level, it tracks population density in the expected direction. The Almnäs case makes the mechanism concrete. Almnäs is a small rural cooperative with total costs of 3,102 kSEK. The LP-solver scales Almnäs up by a factor of 200 to 800 in order to build the frontier benchmark for Sweden's three largest operators: Vattenfall (scale factor 809, 98.2 percent of which falls on Almnäs), E.ON (833, 89.6 percent), and Ellevio (220, 89.2 percent). The LP is doing exactly what it is designed to do: finding the minimum-cost convex combination of observed firms that meets the required output targets. But the resulting comparison has no structural coherence. *Almnäs* operates a rural network serving a fraction of the customers in a fundamentally different terrain, yet it serves as the primary reference point for firms 2,000 times its size in terms of costs. *Almnäs* is itself a LOVO dropper: it loses efficient status when NS is removed. The most widely used peer in the sample is therefore also the firm most dependent on NS for its frontier position.

Table 5.4: Top-5 firms by $\Sigma\lambda$: the Almnäs peer-hub mechanism.

Firm	θ	TOTEX (MSEK)	$\Sigma\lambda$	$\lambda_{\text{Almnäs}}$	Almnäs share
E.ON Energidistr. AB	0.941	6550	833	747	89.6%
Vattenfall Eldistribution AB	0.881	6075	809	795	98.2%
Ellevio AB	0.931	4998	220	197	89.2%
Gotlands Elnät AB	0.737	384	36	33	91.6%
Härjeåns Nät AB	0.814	254	28	27	95.3%
Reference: Almnäs Elnät AB (REL03059) — TOTEX 3.1 MSEK					

Note: $\Sigma\lambda$ is the sum of peer weights under input-oriented CRS DEA, interpretable as the scale factor applied to the efficient peer combination.

The next subsection asks whether this aggregate pattern holds at the firm level, or whether a subset of firms receive NS compensation that their operating environment does not explain.

5.1.4 Firm-level NS–density alignment

Table 5.5: Fractional-logit coefficients of NS share on log density.

	DEA		StoNED	
	Primary	Robust	Primary	Robust
Panel A: Linear specification				
Intercept	2.670^{***}	2.197^{***}	3.252^{***}	2.742^{***}
	(0.352)	(0.519)	(0.510)	(0.643)
$\log(1 + \text{PWD})$	-0.578^{***}	-0.545^{***}	-0.628^{***}	-0.593^{***}
	(0.055)	(0.060)	(0.074)	(0.078)
Snow (max depth, pop-wt.)	—	0.074	—	-0.365
		(0.284)		(0.328)
Forest (%, pop-wt.)	—	0.007	—	0.011^{*}
		(0.006)		(0.007)
Panel B: Quadratic-augmented specification				
Intercept	-2.624^{*}	—	2.171	—
	(1.443)		(2.458)	
$\log(1 + \text{PWD})$	1.238^{**}	—	-0.264	—
	(0.499)		(0.772)	
$[\log(1 + \text{PWD})]^2$	-0.150^{***}	—	-0.030	—
	(0.042)		(0.060)	
N	144	144	144	144

Table 5.5 reports the fractional regression of each firm's NS share on log population-weighted density. The gradient is steep and stable: a one-unit increase in log density reduces the expected NS share by 0.578 in DEA ($z = -10.58$) and 0.628 in StoNED ($z = -8.46$). Adding snow depth and forest cover as controls moves the density coefficient by less than 6 percent in both methods. Of the two environmental variables, only forest cover reaches conventional significance in StoNED ($p < 0.10$); snow and forest are insignificant in DEA, and snow is insignificant in StoNED. The stability of the density coefficient across specifications suggests that NS is primarily tracking population density rather than the broader environmental conditions forest and snow are meant to capture. The aggregate mechanism is working as intended: rural firms receive more NS weight. But the implication is that the environmental detail these additional variables could provide

is largely already absorbed by the density gradient itself. The quadratic term in Panel B is significant for DEA ($p = 0.0004$) but not StoNED ($p = 0.62$), signalling nonlinearity in DEA's NS-density relationship.

Table 5.6: Candidate firms by Pearson residual (over- and under-NS).

	log(1+PWD)	NS share		Pearson z		Owner	Source
		DEA	StoNED	DEA	StoNED		
Over-NS firms (high NS-weight given density; wrong-firm candidates)							
Sturefors Eldistribution AB	6.62	0.50	0.79	+0.62	+1.10	Private	Both
Tekniska verken Katrineholm Nät AB	7.74	0.43	0.43	+0.83	+0.69	Municipal	Both
Vattenfall Eldistribution AB	7.81	0.36	0.45	+0.66	+0.79	State	Both
Ellevio AB	8.60	0.24	0.29	+0.52	+0.60	Private	StoNED
E.ON Energidistribution AB	7.83	0.33	0.37	+0.58	+0.59	Private	Both
Gotlands Elnät AB	6.91	0.43	0.44	+0.54	+0.43	State	DEA
Under-NS firms (low NS-weight given density; missed-firm candidates)							
Näckåns Elnät AB	3.90	0.41	0.41	−0.39	−0.61	Private	StoNED
Kvänumbygdens Energi ek. för.	5.50	0.40	0.15	+0.04	−0.60	Other / coop.	StoNED
Bergs Tingslags Elektriska AB	4.68	0.26	0.28	−0.45	−0.60	Municipal	Both
Olseröds Elektriska Distributionsförening u.p.a.	3.58	0.41	0.49	−0.49	−0.55	Other / coop.	Both
Oxelö Energi AB	7.22	0.10	0.00	−0.21	−0.52	Municipal	StoNED
Filipstad Energinät AB	6.82	0.03	0.11	−0.46	−0.35	Municipal	DEA
Falkenberg Energi AB	7.38	0.00	0.03	−0.44	−0.41	Municipal	DEA
Hallstaviks Elverk ek. för.	7.12	0.02	0.10	−0.44	−0.30	Other / coop.	DEA

Table 5.6 lists these over-NS firms. The three largest Swedish DSOs, Vattenfall (log density 7.81, NS share 0.36 in DEA), E.ON (7.83, 0.33), and Ellevio (8.60, 0.24), all sit above the predicted band despite operating predominantly urban networks. The pattern is more pronounced in StoNED, where Sturefors reaches an NS share of 0.79 at a density of 6.62, and Vattenfall reaches 0.45.

These over-NS firms are large operators with geographically dispersed concession areas that include both dense urban centres and sparser peripheral zones. The NS count reflects the total network footprint, but the density measure reflects the population-weighted average. The gap between the two is where the compensation mechanism breaks down: firms receive structural credit for their total network span while their average operating environment is urban.

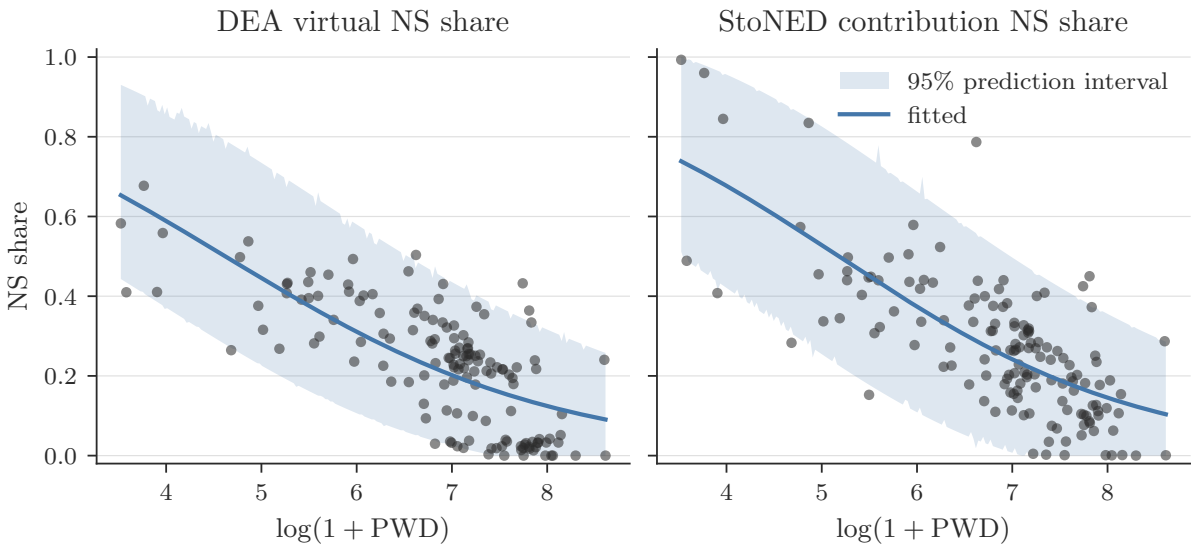


Figure 5.2: Predicted NS share against log population-weighted density: DEA (left) and StoNED (right). Shaded band shows 95% prediction interval from fractional regression.

Figure 5.2 plots each firm's NS share against its density, with the fitted curve and 95 percent prediction band from the fractional regression. Firms with large positive Pearson residuals from the fractional regression are receiving more NS compensation than their density predicts. Table 5.6 from the previous paragraph lists these candidates by name and Figure 5.2 shows the aggregate spread of the observations.

5.2 The Parametric z-Channel as Explicit Compensation (RQ2)

5.2.1 Delta-hat inference

Table 5.7: $\hat{\delta}$ from StoNEZD second-stage OLS, single-z and full-z.

Variable	$\hat{\delta}$	Robust SE	t	p	95% CI	Partial R^2
Single-z specs						
Linear pop dens. ($\times 10^{-4}$)	−1.472	0.234	−6.29	< 0.001	[−1.93, −1.01]	0.218
Log pop dens.	−0.162	0.019	−8.53	< 0.001	[−0.20, −0.12]	0.339
Arith. density ($\times 10^{-4}$)	−4.446	0.636	−6.99	< 0.001	[−5.69, −3.20]	0.256
Polyagg area ($\times 10^{-4}$)	−1.655	0.242	−6.83	< 0.001	[−2.13, −1.18]	0.247
Full-z anchor (popwt)						
Linear PWD ($\times 10^{-4}$)	−1.296	0.193	−6.72	< 0.001	[−1.67, −0.92]	0.244
max depth	0.270	0.299	0.90	0.367	[−0.32, +0.86]	0.006
forest % ($\times 10^{-3}$)	1.998	1.920	1.04	0.298	[−1.76, +5.76]	0.008

Table 5.7 reports the estimated z-coefficient $\hat{\delta}$ from the StoNEZD second-stage OLS, where the dependent variable is the log-deviation of actual TOTEX from the CNLS output composite $\hat{\boldsymbol{\beta}}'\mathbf{y}$. In the multiplicative CRS specification, this residual isolates $\delta \cdot z + \varepsilon$, so regressing it on z recovers $\hat{\delta}$ directly. The upper panel varies the density measure across four single-z specifications:

the population-weighted density (linear PWD), its log transform, arithmetic mean density, and polygon-area-weighted density. The lower panel admits snow depth and forest cover alongside PWD in a full-z specification.

In each of the four single-z specs, $\hat{\delta}$ is negative and statistically significant ($p < 0.001$), with partial R^2 of 0.218 for linear PWD and 0.339 for log PWD. The residual log-costs are lower in denser concession areas, as expected if density captures cost-relevant heterogeneity that the output composite does not. The magnitude varies with functional form and aggregation method, as expected given the different scalings, but the direction and significance are stable across all four specifications.

Adding snow depth and forest cover alongside PWD reduces delta-hat by about 12 percent relative to the single-z headline (-1.296×10^{-4} vs -1.472×10^{-4}), while partial R^2 rises slightly from 0.218 to 0.244. Snow depth and forest cover are not statistically significant. Including additional environmental variables, therefore, does not weaken the density channel; PWD is the only z-variable that carries a detectable cost shift, and the downstream cohort falsification test rests on this result.

5.2.2 Cohort falsification

The $\hat{\delta}$ estimates establish that PWD shifts log-costs at the sample level. The cohort falsification tests whether this shift corrects the specific firms RQ1 identified as miscompensated by NS. The micro diagnostic flagged two groups: wrong-firm candidates (over-NS), whose NS share exceeded what their density predicts, and missed-firm candidates (under-NS), whose NS share fell short. If the z-channel resolves NS-mediated heterogeneity, wrong-firm candidates should see both lower cost positions and lower efficiency ratings under StoNEZD relative to the NS-inclusive k=5 baseline, while missed-firm candidates should see higher efficiency ratings.

Figure 5.3 plots ΔC against ΔT for all 144 firms; the two groups are highlighted, and their centroids are indicated. The means and standard deviations are reported in Table 5.8. The wrong-firm group lands in the predicted SW quadrant: mean $\Delta C = -0.510$ and mean $\Delta T = -0.213$. Switching from NS to the z-channel lowers both their cost position and their measured efficiency, consistent with these firms having received excess compensation under the k=5 specification.

Table 5.8: ΔC, ΔT centroids by cohort.

Cohort	n	$\overline{\Delta C}$	SD ΔC	$\overline{\Delta T}$	SD ΔT	Predicted	Actual	Match
Wrong-firm (over-NS)	5	−0.510	0.168	−0.213	0.074	SW	SW	✓
Missed-firm (under-NS)	5	−0.118	0.165	+0.023	0.052	NE	NW	×

The missed-firm group lands in the NW quadrant rather than the predicted NE. Mean ΔT is positive, confirming that StoNEZD raises its efficiency rating as expected. However, mean ΔC is negative: their costs still fall under StoNEZD because NS gave them some absolute cost relief in the k=5 model that the z-channel does not fully offset. The efficiency prediction holds, but the cost prediction does not. The asymmetry reflects the difference between relative and absolute

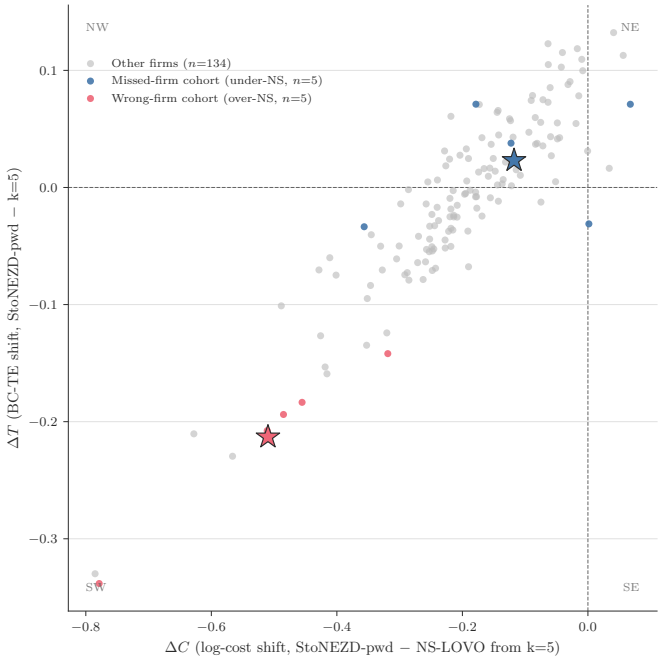


Figure 5.3: Firm-level ΔC vs. ΔT scatter; cohorts highlighted with centroids.

compensation. The micro diagnostic identifies under-NS firms relative to the density gradient, but the absolute cost credit NS provided in the $k = 5$ model exceeds what $\hat{\delta} \cdot \text{PWD}$ recovers.

5.2.3 MAUP sensitivity

The previous subsections use population-weighted density as the z-variable. Because concession areas vary in shape and population distribution, the way density is aggregated to the firm level is itself a modelling choice. Table 5.9 tests whether the results for the three largest DSOs, Vattenfall, E.ON, and Ellevio, survive alternative aggregation methods. The single-z panel varies the measure across four specifications (linear PWD, log PWD, arithmetic density, polygon-area-weighted). The full-z panel holds PWD as the anchor and adds snow depth and forest cover under three aggregation variants (population-weighted, area-weighted, snowdays). Each cell reports the shift in log-cost (ΔC) and the shift in BC technical efficiency (ΔT).

Table 5.9: MAUP sensitivity for the Big Three: ΔC, ΔT.

Panel A: single-z aggregations

	pop dens		log(pop dens)		arith. dens.		polyagg area	
Firm	ΔC	ΔT	ΔC	ΔT	ΔC	ΔT	ΔC	ΔT
Vattenfall	−0.511	−0.208	−1.413	−0.198	−0.152	−0.079	−0.253	−0.036
E.ON	−0.455	−0.183	−1.353	−0.163	−0.095	−0.046	−0.160	−0.002
Ellevio	−0.779	−0.338	−1.372	−0.157	0.007	0.024	−0.055	0.062

Panel B: full-z aggregations

	popwt		areawt		snowdays	
Firm	ΔC	ΔT	ΔC	ΔT	ΔC	ΔT
Vattenfall	−0.313	−0.166	−0.231	−0.067	−0.351	−0.163
E.ON	−0.289	−0.163	−0.325	−0.143	−0.325	−0.160
Ellevio	−0.549	−0.283	−0.619	−0.279	−0.588	−0.279

The direction of ΔC is stable: all three firms show negative cost shifts across every specification in both panels. The direction of ΔT is also negative throughout the single-z panel and for all firms under the population-weighted and area-weighted full-z specs. The headline finding, that the z-channel lowers both cost positions and efficiency ratings for the Big Three, is therefore not sensitive to how density is constructed.

6 Discussion

6.1 NS reliance and the limits of implicit environmental compensation

Across both methods, NS is the single most influential output variable; removing it changes efficiency scores more than dropping any other output. At the aggregate level, NS compensation works as intended: rural firms with more stations receive more credit, and the correlation with population density is strong and stable across both methods. However, the mechanism breaks down at the firm level. Station count captures the total network footprint but does not distinguish between a firm that has many stations because it serves a genuinely rural area and one that has many stations simply because it is large and geographically spread out. The three largest DSOs all receive substantial NS credit despite serving predominantly urban customer bases. Their high station counts reflect geographic spread, not rural hardship. If the model overcompensates these firms, their revenue caps are set too generously, and their customers pay more than they should.

6.2 The curse of dimensionality is a curse of NS

Chapter 3 left one question open. On the high-voltage subsample, the orthogonal residual of MWh_h ($1 - R^2 = 0.254$) is larger than that of NS (0.205), and dropping MWh_h reduces the output cloud's effective rank by more than dropping NS does (0.061 versus 0.032; Appendix Table A.4). On a naive pure geometric reading, MWh_h reads as the stronger candidate for a second axis, orthogonal to the scale dimension, that the frontier estimation could exploit as an escape direction. However, the macro level (Table 5.1) and LOVO results (Table 5.2) all point to NS being the most influential variable for the frontier in both methods. The apparent inconsistency echoes the lesson from previous analysis: an aggregate measure can point one way while the firm-level distribution beneath it points another.

We first observe that the distributional characteristics of our two residuals have opposite shape. The residual distribution of NS is near-Normal and symmetric with excess kurtosis -0.34, a largest single residual of only 1.48 standardised units, and the five most extreme firms holding just 10.4% of the total residual mass. The implication is that each firm carries a non-trivial NS residual. The distribution of MWh_h's residual is the opposite with a heavy-tail and strong left skew (excess kurtosis 4.4, largest residual 4.28), with the top five firms holding 17.6% of total residual mass (Appendix Table A.3). Its apparent orthogonality from the pure geometric reading is therefore manufactured by a few firms with unusual high-voltage profile relative their other operations; macro level analysis of effective rank of the output cloud should therefore also include a micro-level analysis of how the orthogonal mass is distributed.

The difference of the orthogonal mass propagates directly to their importance to the frontier,

and the frontier estimates are decisive. We won't repeat results previously discussed, but the important takeaway is that a variable with non-trivial orthogonal residual for all firms will determine the shape of the frontier to a much larger degree than the other variables. On a nominal count, the five outputs would probably flag the specification for overfitting; with the effective rank refinement, however, we show that the intrinsic dimension of the output cloud is the relevant measure and NS is the lone direction which a firm can escape its disciplining neighbours. In the Swedish specification, the curse of dimensionality is a curse of NS.

A case study in its implication can be found in Almnäs. The firm runs a small rural network alongside a farm with conference facilities, cheese production, and a single wind turbine, from which it consumes roughly 40% of the output itself; the rest of the network serves vacation homes and small villages, at total costs of 3,102 kSEK. It is also a firm whose frontier position is defined by NS, sitting at the frontier in DEA but sees its efficiency score reduce by 0.47 (DEA) and 0.346 (StoNED) points when NS is removed from the specification. Its efficiency rests entirely on the one orthogonal axis along which few disciplinary neighbours sit in the output cloud to constrain it. Constant returns to scale then turns this local quirk into a sample-wide one. Because the CRS technology is a cone, any efficient point can be scaled to any size and stay on the frontier, so the linear program scales Almnäs up by a factor of 809 to build 98.2% of Vattenfall's benchmark, and does much the same for E.ON and Ellevio (Table 5.4). The firm whose own efficiency is least disciplined thereby becomes the benchmark for Sweden's three largest DSOs and, read literally, recommends they book a conference at the cheese farm to study best practice.

6.3 Can StoNED deliver what DEA cannot?

StoNEZD's z-channel corrects the most prominent problem that RQ1 identifies. The firms that received more credit for their station count than their operating environment would suggest, mostly the large urban operators, see both their cost position and efficiency rating fall when density is handled through the z-channel instead. The correction holds regardless of how density is measured and regardless of whether snow and forest are included as additional environmental controls. The key difference is in how the adjustment is made. When NS sits in the output vector, the amount of density compensation each firm receives is a byproduct of the model's weight selection, different for every firm and not subject to any statistical test. When density enters through the z-channel, the adjustment is a single estimated coefficient applied uniformly across all firms, and its magnitude, significance, and confidence interval are all reported. The regulator can see exactly how large the density correction is and whether it is statistically distinguishable from zero.

The correction does not work equally well in both directions, however. Firms that were under-compensated by NS see their efficiency rating improve, as expected, but their overall cost position does not recover. The reason comes down to what NS actually does inside the model. NS plays two roles at once: it captures genuine production costs (more stations means more transformers to maintain and replace) and it acts as a proxy for how rural or urban a firm's operating environment is. When density is moved to the z-channel, only the second role is replaced. The first role, the

real cost of building and maintaining stations, disappears from the model along with the variable. For firms that genuinely need many stations to serve their customers, this means the model no longer gives them full credit for a real cost they bear. A regulator considering this trade-off would need to weigh the gain in transparency from separating the two roles against the loss in cost recognition for firms where station count reflects genuine infrastructure needs rather than geographic spread.

6.4 Implications for RP5

These results do not point to a simple recommendation of one method over another, but they do clarify what is at stake for the RP5 design. Under RP5, TOTEX benchmarking makes the efficiency deduction the primary lever on both operating costs and capital investment, so the blending of production efficiency with environmental compensation identified above matters more than it did under the current framework.

StoNEZD offers a feasible way to separate the two roles. The z-channel produces a transparent, testable density adjustment that corrects efficiency scores for the firms where NS compensation is most misaligned, and the correction is robust to how density is measured. Adopting it would, however, require two explicit design choices: whether to include snow and forest alongside density (neither is statistically significant in this sample, but a regulator may want them as a precaution), and how to aggregate density for firms whose concession areas span multiple disconnected regions.

7 Conclusion

Sweden's energy regulator relies on network stations to give firms credit for operating in difficult environments. At the aggregate level, this works: rural firms with more stations receive more compensation. But at the firm level, the mechanism breaks down. The three largest operators receive substantial NS credits not because they serve rural areas, but because their networks span multiple regions. Their customers pay more than they need to simply because the mechanism can't distinguish between firms that need the credit and firms that simply happen to be large.

StoNEZD's z-channel fixes the clearest part of this problem. It strips excess compensation from the firms that do not need it and produces a density adjustment that the regulator can inspect, test, and defend. The fix is robust to how density is measured. It is not complete, though. Firms that genuinely depend on many stations lose some cost recognition when NS is removed.

References

Adler, Nicole and Boaz Golany (2001). "Evaluation of deregulated airline networks using data envelopment analysis combined with principal component analysis with an application to Western Europe". In: *European Journal of Operational Research* 132.2, pp. 260–273.

– (2002). "Including principal component weights to improve discrimination in data envelopment analysis". In: *Journal of the Operational Research Society* 53.9, pp. 985–991.

Adler, Nicole and Ekaterina Yazhemsky (2010). "Improving discrimination in data envelopment analysis: PCA–DEA or variable reduction". In: *European Journal of Operational Research* 202.1, pp. 273–284.

Afriat, Sidney N. (1972). "Efficiency estimation of production functions". In: *International Economic Review*, pp. 568–598.

Aigner, Dennis, C. A. Knox Lovell, and Peter Schmidt (1977). "Formulation and estimation of stochastic frontier production function models". In: *Journal of Econometrics* 6.1, pp. 21–37.

Allen, Rachel, Antreas Athanassopoulos, Robert G. Dyson, and Emmanuel Thanassoulis (1997). "Weights restrictions and value judgements in data envelopment analysis: evolution, development and future directions". In: *Annals of Operations Research* 73.0, pp. 13–34.

Arocena, Pablo, Subal C. Kumbhakar, and Gudbrand Lien (2025). "Performance benchmarking for regulating electricity distribution: Insights from empirical academic research". In: *Journal of Economic Surveys* 39.3, pp. 1287–1324.

Banker, Rajiv, Ram Natarajan, and Daqun Zhang (2019). "Two-stage estimation of the impact of contextual variables in stochastic frontier production function models using data envelopment analysis: Second stage OLS versus bootstrap approaches". In: *European Journal of Operational Research* 278.2, pp. 368–384.

Bárány, Imre (2008). "Random points and lattice points in convex bodies". In: *Bulletin of the American Mathematical Society* 45.3, pp. 339–365.

Bárány, Imre and David G Larman (1988). "Convex bodies, economic cap coverings, random polytopes". In: *Mathematika* 35.2, pp. 274–291.

Battese, George E. and Tim J. Coelli (1988). "Prediction of firm-level technical efficiencies with a generalized frontier production function and panel data". In: *Journal of Econometrics* 38.3, pp. 387–399.

Bertsimas, Dimitris and Nishanth Mundru (2021). "Sparse convex regression". In: *INFORMS Journal on Computing* 33.1, pp. 262–279.

Bjørndal, Endre, Mette Bjørndal, Astrid Cullmann, and Maria Nieswand (2016). *Finding the right yardstick: Regulation under heterogeneous environments*. Tech. rep. DIW Discussion Papers.

Bogetoft, Peter (1997). "DEA-based yardstick competition: The optimality of best practice regulation". In: *Annals of Operations Research* 73.0, pp. 277–298.

Charnes, Abraham, William W. Cooper, and Edwardo Rhodes (1978). "Measuring the efficiency of decision making units". In: *European Journal of Operational Research* 2.6, pp. 429–444.

Cook, Wade D. and Larry M. Seiford (2009). "Data envelopment analysis (DEA)–Thirty years on". In: *European Journal of Operational Research* 192.1, pp. 1–17.

Cordero, José Manuel, Cristina Polo, and Daniel Santín (2020). "Assessment of new methods for incorporating contextual variables into efficiency measures: a Monte Carlo simulation". In: *Operational Research* 20.4, pp. 2245–2265.

Craig, John (1984). "Averaging population density". In: *Demography* 21.3, pp. 405–412.

Daraio, Cinzia and Leopold Simar (2007). *Advanced robust and nonparametric methods in efficiency analysis: Methodology and applications*. Springer.

Daraio, Cinzia and Léopold Simar (2005). "Introducing environmental variables in nonparametric frontier models: a probabilistic approach". In: *Journal of productivity analysis* 24.1, pp. 93–121.

Daraio, Cinzia, Léopold Simar, and Paul W Wilson (2018). "Central limit theorems for conditional efficiency measures and tests of the 'separability'condition in non-parametric, two-stage models of production". In: *The Econometrics Journal* 21.2, pp. 170–191.

Doyle, John and Rodney Green (1994). "Efficiency and cross-efficiency in DEA: Derivations, meanings and uses". In: *Journal of the Operational Research Society* 45.5, pp. 567–578.

Dyson, Robert G., Rachel Allen, Ana S. Camanho, Victor V. Podinovski, Cláudia S. Sarrico, and Estelle A. Shale (2001). "Pitfalls and protocols in DEA". In: *European Journal of Operational Research* 132.2, pp. 245–259.

Dyson, Robert G. and Emmanuel Thanassoulis (1988). "Reducing weight flexibility in data envelopment analysis". In: *Journal of the Operational Research Society* 39.6, pp. 563–576.

Edvardsen, Dag Føsund, Finn R. Førsund, and Sverre A. C. Kittelsen (2008). "Far out or alone in the crowd: a taxonomy of peers in DEA". In: *Journal of Productivity Analysis* 29.3, pp. 201–210.

Energiavirasto (2023a). *Explanatory memorandum: Confirmation decisions regulation methods in the sixth regulatory period from 1 January 2024 to 31 December 2027 and for the seventh regulatory period from 1 January 2028 to 31 December 2031 – electricity distribution network operations and high-voltage distribution network operations*. Appendix 3 to confirmation decision Dnro 3171/040300/2023. Energiavirasto [Finnish Energy Authority].

– (2023b). *Regulation methods in the sixth regulatory period of 1 January 2024 – 31 December 2027 and the seventh regulatory period of 1 January 2028 – 31 December 2031 – electricity distribution network operations and high-voltage distribution network operations*. Appendix 2 to confirmation decision Dnro 3171/040300/2023. Energiavirasto [Finnish Energy Authority].

Energimarknadsinspektionen (2024). *Effektiviseringskrav för lokalnätsföretag: För tillsynsperioden 2024–2027 [Efficiency requirements for local grid companies: For the regulatory period 2024–2027]*. Bilaga 9. Energimarknadsinspektionen.

Fan, Yanqin, Qi Li, and Alfons Weersink (1996). "Semiparametric estimation of stochastic production frontier models". In: *Journal of Business & Economic Statistics* 14.4, pp. 460–468.

Fotheringham, A. Stewart and David W. S. Wong (1991). "The modifiable areal unit problem in multivariate statistical analysis". In: *Environment and Planning A* 23.7, pp. 1025–1044.

Ghosal, Promit and Bodhisattva Sen (2017). "On univariate convex regression". In: *Sankhya A* 79.2, pp. 215–253.

Han, Qiyang and Jon A Wellner (2016). "Multivariate convex regression: global risk bounds and adaptation". In: *arXiv preprint arXiv:1601.06844*.

Hildreth, Clifford (1954). "Point estimates of ordinates of concave functions". In: *Journal of the American Statistical Association* 49.267, pp. 598–619.

Jenkins, Larry and Murray Anderson (2003). "A multivariate statistical approach to reducing the number of variables in data envelopment analysis". In: *European Journal of Operational Research* 147.1, pp. 51–61.

Johnson, Andrew L. and Timo Kuosmanen (2011). "One-stage estimation of the effects of operational conditions and practices on productive performance: asymptotically normal and efficient, root-n consistent StoNEZD method". In: *Journal of Productivity Analysis* 36.2, pp. 219–230.

– (2012). "One-stage and two-stage DEA estimation of the effects of contextual variables". In: *European Journal of Operational Research* 220.2, pp. 559–570.

Jondrow, James, C. A. Knox Lovell, Ivan S. Materov, and Peter Schmidt (1982). "On the estimation of technical inefficiency in the stochastic frontier production function model". In: *Journal of Econometrics* 19.2-3, pp. 233–238.

Joskow, Paul L. (2007). "Regulation of natural monopoly". In: *Handbook of Law and Economics* 2, pp. 1227–1348.

– (2008). "Lessons learned from electricity market liberalization". In: *The Energy Journal* 29.2_suppl, pp. 9–42.

Kneip, Alois, Byeong U. Park, and Léopold Simar (1998). "A note on the convergence of nonparametric DEA estimators for production efficiency scores". In: *Econometric Theory* 14.6, pp. 783–793.

Koltchinskii, Vladimir and Karim Lounici (2017). "Concentration inequalities and moment bounds for sample covariance operators". In: *Bernoulli* 23.1, pp. 110–133.

Kordahl, Ole-Petter, Hilde Marit Kvile, and Fartein Valen Slåttebrekk (2022). *Endringer i modell for kostnadsnorm. Korrigering for rammevilkår: Oppsummering av høringsinnspill og endelig modell [Changes to the cost-norm model. Correction for framework conditions: Summary of consultation input and final model]*. RME Rapport 10/2022. Reguleringsmyndigheten for energi (RME), Norges vassdrags- og energidirektorat (NVE).

Korostelev, Alexander, Léopold Simar, and A Tsybakov (1995). "On estimation of monotone and convex boundaries". In: *Annales de l'ISUP*. Vol. 39. 1, pp. 3–18.

Kuosmanen, Timo (2008). "Representation theorem for convex nonparametric least squares". In: *The Econometrics Journal* 11.2, pp. 308–325.

– (2018). "Conditional yardstick competition". In: *Aalto University*.

Kuosmanen, Timo and Andrew L. Johnson (2010). "Data envelopment analysis as nonparametric least-squares regression". In: *Operations Research* 58.1, pp. 149–160.

– (2017). "Modeling joint production of multiple outputs in StoNED: Directional distance function approach". In: *European Journal of Operational Research* 262.2, pp. 792–801.

– (2021). "Conditional yardstick competition in energy regulation". In: *The Energy Journal* 42.1_suppl, pp. 1–26.

Kuosmanen, Timo and Mika Kortelainen (2012). "Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints". In: *Journal of Productivity Analysis* 38.1, pp. 11–28.

Kur, Gil, Fuchang Gao, Adityanand Guntuboyina, and Bodhisattva Sen (2020). "Convex regression in multidimensions: Suboptimality of least squares estimators". In: *arXiv preprint arXiv:2006.02044*.

Laffont, Jean-Jacques and Jean Tirole (1993). *A theory of incentives in procurement and regulation*. Cambridge, MA: MIT Press.

Lei, Jing, Max G'Sell, Alessandro Rinaldo, Ryan J. Tibshirani, and Larry Wasserman (2018). "Distribution-free predictive inference for regression". In: *Journal of the American Statistical Association* 113.523, pp. 1094–1111.

Liao, Zhiqiang, Sheng Dai, Eunji Lim, and Timo Kuosmanen (2024). "Overfitting reduction in convex regression". In: *arXiv preprint arXiv:2404.09528*.

– (2026). "Boundary problem and overfitting reduction in convex regression". In: *European Journal of Operational Research*.

Lim, Eunji (2021). "Consistency of penalized convex regression". In: *International Journal of Statistics and Probability* 10.1.

Lim, Eunji and Peter W. Glynn (2012). "Consistency of multidimensional convex regression". In: *Operations Research* 60.1, pp. 196–208.

Lundin, Erik and Magnus Söderberg (2022). *Analys av priser och reglering på den svenska elnätsmarknaden: Vad kan vi lära av våra grannländer?* Stockholm: SNS Förlag.

Mazumder, Rahul, Arkopal Choudhury, Garud Iyengar, and Bodhisattva Sen (2019). "A computational framework for multivariate convex regression and its variants". In: *Journal of the American Statistical Association* 114.525, pp. 318–331.

Meeusen, Wim and Julien van Den Broeck (1977). "Efficiency estimation from Cobb-Douglas production functions with composed error". In: *International Economic Review*, pp. 435–444.

Nieswand, Maria and Stefan Seifert (2018). "Environmental factors in frontier estimation–A Monte Carlo analysis". In: *European Journal of Operational Research* 265.1, pp. 133–148.

Openshaw, Stan (1984). *The modifiable areal unit problem*. Norwich: GeoBooks.

Ottensmann, John R. (2018). "On population-weighted density". In: *SSRN Electronic Journal*.

Papke, Leslie E. and Jeffrey M. Wooldridge (1996). "Econometric methods for fractional response variables with an application to 401 (k) plan participation rates". In: *Journal of Applied Econometrics* 11.6, pp. 619–632.

Park, Byeong U., Seok-Oh Jeong, and Léopold Simar (2010). "Asymptotic distribution of conical-hull estimators of directional edges". In.

Podinovski, Victor V. (2004). "Production trade-offs and weight restrictions in data envelopment analysis". In: *Journal of the Operational Research Society* 55.12, pp. 1311–1322.

– (2005). "The explicit role of weight bounds in models of data envelopment analysis". In: *Journal of the Operational Research Society* 56.12, pp. 1408–1418.

Ray, Subhash C. (1988). "Data envelopment analysis, nondiscretionary inputs and efficiency: an alternative interpretation". In: *Socio-Economic Planning Sciences* 22.4, pp. 167–176.

Ruggiero, John (1998). "Non-discretionary inputs in data envelopment analysis". In: *European Journal of Operational Research* 111.3, pp. 461–469.

Saastamoinen, Antti and Timo Kuosmanen (2016). "Quality frontier of electricity distribution: Supply security, best practices, and underground cabling in Finland". In: *Energy Economics* 53, pp. 281–292.

Sarrico, Cláudia S. and Robert G. Dyson (2004). "Restricting virtual weights in data envelopment analysis". In: *European Journal of Operational Research* 159.1, pp. 17–34.

Seijo, Emilio and Bodhisattva Sen (2011). "Nonparametric least squares estimation of a multivariate convex regression function".

Sexton, Thomas R., Richard H. Silkman, and Andrew J. Hogan (1986). "Data envelopment analysis: Critique and extensions". In: *New Directions for Program Evaluation* 1986.32, pp. 73–105.

Shleifer, Andrei (1985). "A theory of yardstick competition". In: *The RAND Journal of Economics*, pp. 319–327.

Simar, Léopold, Anne Vanhems, and Ingrid Van Keilegom (2016). "Unobserved heterogeneity and endogeneity in nonparametric frontier estimation". In: *Journal of Econometrics* 190.2, pp. 360–373.

Simar, Léopold and Paul W. Wilson (2007). "Estimation and inference in two-stage, semi-parametric models of production processes". In: *Journal of Econometrics* 136.1, pp. 31–64.

Söderberg, Magnus and Mattias Vesterberg (2025). "Scale properties and efficient network structures in the Swedish electricity distribution market". In: *Journal of Regulatory Economics* 68.1, pp. 17–37.

Stone, Charles J. (1980). "Optimal rates of convergence for nonparametric estimators". In: *The Annals of Statistics*, pp. 1348–1360.

Thanassoulis, Emmanuel, Maria C. S. Portela, and Ozren Despic (2008). "Data envelopment analysis: the mathematical programming approach to efficiency analysis". In: *The Measurement of Productive Efficiency and Productivity Growth* 1, pp. 251–420.

Torgersen, Arne M., Finn R. Førsund, and Sverre A. C. Kittelsen (1996). "Slack-adjusted efficiency measures and ranking of efficient units". In: *Journal of Productivity Analysis* 7.4, pp. 379–398.

Vershynin, Roman (2026). *High-Dimensional Probability: An Introduction with Applications in Data Science*. 2nd ed. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press.

Wilson, Paul W. (2018). "Dimension reduction in nonparametric models of production". In: *European Journal of Operational Research* 267.1, pp. 349–367.

Wong, Yue-Hong B. and John E. Beasley (1990). "Restricting weight flexibility in data envelopment analysis". In: *Journal of the Operational Research Society* 41.9, pp. 829–835.

A Supplementary Material

A.1 Best practice versus average practice frontier

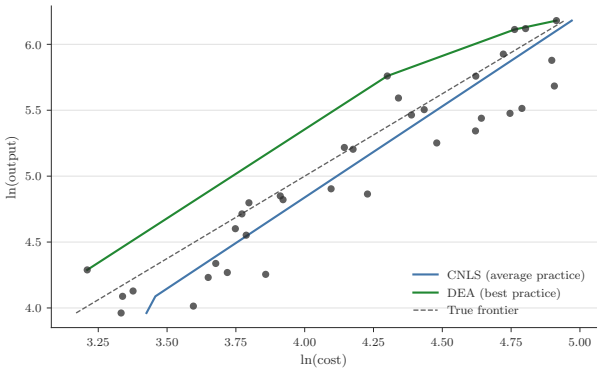


Figure A.1: Best practice versus average practice frontier.

A.2 Concession dispersion

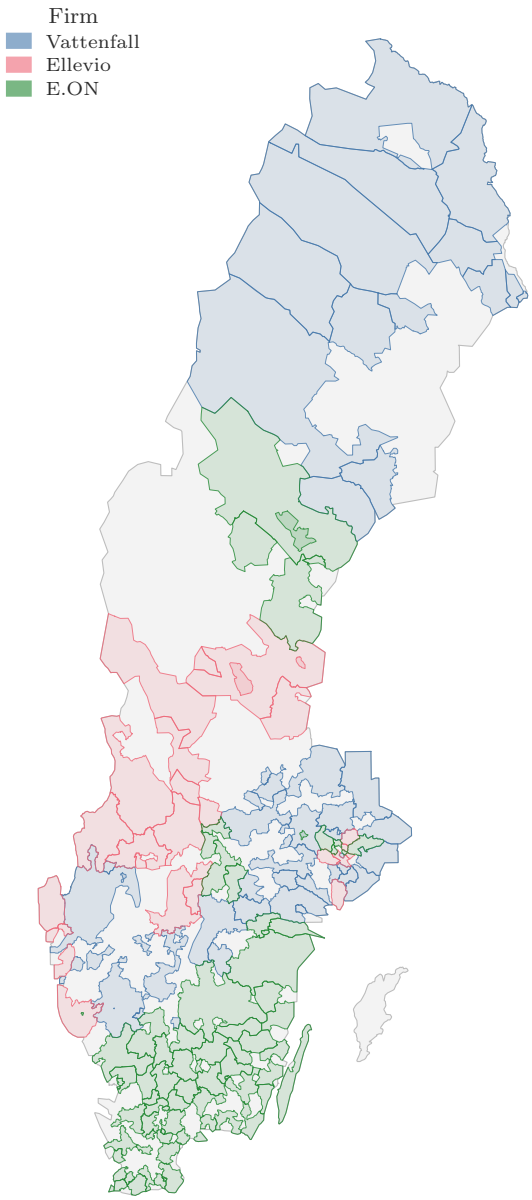


Figure A.2: Concession dispersion of the three largest Swedish DSOs. Each firm's service area is the union of geographically disjoint concession polygons spread across Sweden, illustrating that a single firm-level centroid is a poor summary of where a firm actually operates.

A.3 Output cloud orthogonality

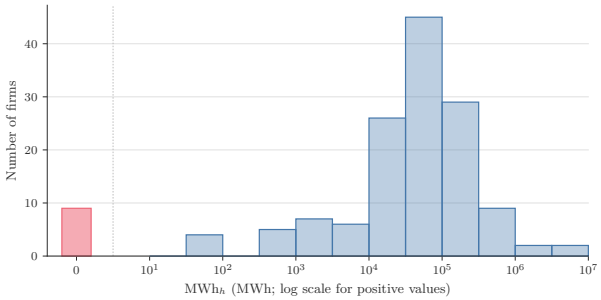


Figure A.3: Distribution of MWh_h across the 144 Swedish DSOs. Nine firms report exactly zero; the threshold is an exact-zero discontinuity rather than a low-density left tail.

Table A.1: $1 - R^2$ from OLS of each log-output on the other four log-outputs. HV operator sub-sample, $n = 135$.

Dependent variable	R^2	$1 - R^2$
log CU	0.981	0.019
log MW	0.975	0.025
log NS	0.795	0.205
log MWh_ℓ	0.990	0.010
log MWh_h	0.746	0.254

Table A.2: Two-part decomposition of MWh_h. The categorical part (Panel A) uses the pooled sample of 144 firms; the continuous part (Panel B) uses the HV sub-sample of 135 firms and replicates the MWh_h row of Table A.1.

Panel A: Probit of $\mathbf{1}\{MWh_h > 0\}$ *on* $\{\log CU, \log MW, \log NS, \log MWh_\ell\}$*, pooled* $n = 144$	
McFadden pseudo-R^2	0.343
Classification accuracy	0.951
Panel B: OLS of $\log MWh_h$ *on* $\{\log CU, \log MW, \log NS, \log MWh_\ell\}$*, HV sub-sample* $n = 135$	
R^2	0.746
$1 - R^2$	0.254

Table A.3: Distribution of orthogonal residuals (high-voltage sub-sample, $n = 135$).

Panel A: Distributional diagnostics of orthogonal OLS residuals ($n = 135$)				
Residual of	Kurt.	Skew.	Max $\lvert e \rvert$	Top-5
log CU	3.61	−0.86	0.72	0.159
log MW	5.09	1.73	0.88	0.181
log NS	−0.34	−0.14	1.48	0.104
log MWh_ℓ	2.26	0.30	0.52	0.142
log MWh_h	4.39	−1.80	4.28	0.176
Panel B: Percentile bootstrap CI for NS $1 - R^2$*, 2,000 replicates, seed 20260514*				
Point estimate		0.205		
Bootstrap mean (SE)		0.204 (0.042)		
95% CI		[0.132, 0.295]		

Table A.4: Effective rank $r_e(\Sigma) = \mathrm{tr}(\Sigma)/\|\Sigma\|_{\mathrm{op}}$ of the standardised log-output covariance, full five-output cloud and each four-output cloud obtained by dropping one variable. $\Delta r_e = r_e^{\mathrm{drop}} - r_e^{(5)}$; negative entries indicate the dropped variable lay on a minor axis. HV sub-sample, $n = 135$.

Cloud	r_e	Δr_e
Five-output (all)	1.119	—
Drop log CU	1.134	+0.015
Drop log MW	1.143	+0.025
Drop log NS	1.087	−0.032
Drop log MWh_ℓ	1.140	+0.021
Drop log MWh_h	1.058	−0.061

A.4 Revision report

Overall structure

- The two former results chapters (Chapters 5 and 6) were merged into a single Results chapter.
- The conclusion was moved out of the Discussion into a standalone chapter.
- The appendix was expanded.

Chapter 1: Introduction

- Rewritten to reflect a more focused thesis.
- Added a new contributions paragraph.

Chapter 2: Theory and Literature Review

- completely rewritten

Chapter 3: Data

- Added an analysis of output cloud geometry (Section 3.2).

Chapter 4: Methodology

- General grammar edits to make the chapter easier to read.
- Removed the peer-coherence diagnostic, as it was not sufficiently robust.

DEA section.

- The envelopment LP is now written out in full.

- Intuitive interpretations of the constraints added.
- Multiplier (dual) formulation added: output weights, virtual outputs, and virtual output shares defined and explained.

StoNED section.

- The expected-inefficiency formula is explained.
- The direction of u in the cost orientation is stated at the beginning of the introduction.
- The Battese–Coelli TE estimator is justified and explained.

StoNEZD section.

- Clarified that "multiplicative" refers to the error structure, not to how z enters.
- Kuosmanen's OLS-equivalence result is stated explicitly and given its own equation.

Diagnostics section.

- Replaced the "x/z-block" terminology with "output/environmental component".
- Specified the type of standard errors in more detail (HC3).
- MAUP panels renamed to "Panel A" and "Panel B".
- Corrected several typos.
- Wrong-firm compensation and missed-firm compensation are now introduced in the methodology.
- Peer coherence removed completely.
- Changed the order of the diagnostics.

Chapter 5: Results

Removed entirely:

- The D2 diagnostic (Afriat-slack / peer coherence).
- The peer-coherence migration section connected to D2.

Section structure:

- Unified the two results chapters into one.
- Extensive grammar and language improvements to make the chapter more digestible.

Chapter 6: Discussion

- Peer coherence removed and replaced by the Almnäs example.
- The conclusion moved to its own chapter.
- The "Implications for RP5" section trimmed down.

Chapter 7: Conclusion

- Rewritten as a standalone chapter, covering: the aggregate NS mechanism works; the firm-level breakdown for the three largest DSOs; StoNEZD corrects the most prominent misalignment; the correction is incomplete for genuinely rural firms; and the RP5 stakes.

Chapter 8: Supplementary Material

- The previous appendix contained only the AI disclosure; several additions were made.