mathematics

MDPI

Article

Stage Diagnosis in Two-Stage Network SBM: Optimal-Face Bounds and Specification Sensitivity

Guohui Huang [1], Chenjian Shi [2], Jingli Huang [2] and Jie Huang [1,2,*]

[1] Science and Technology Development Department, Fuzhou University, Fuzhou 350108, China
[2] College of Electrical Engineering and Automation, Fuzhou University, Fuzhou 350108, China
* Correspondence: jie.huang@fzu.edu.cn

Abstract

In linked two-stage data envelopment analysis, identifying the weaker stage depends on the modeling choices defining the optimization problem. We examine this dependence for a system in which research resources produce granted invention patents, and patents produce transfer contracts and income, using data for 28 Chinese provinces from 2018 to 2023. We compare output-oriented radial, independent slacks-based, and network slacks-based programs, nine lag alignments and alternative reference sets. Because the optimal system value need not determine unique stage efficiencies, we bound them over the system-optimal face with two auxiliary linear programs; an affine relation between the stage expansions makes these bounds exact, and we state when they are non-trivial. Commercialization efficiencies change by mean absolute differences of 0.104–0.120 at each production-specification step, and technology-formation efficiencies by at most 0.012. Under the two principal alignments, commercialization is classified as weaker in 63–65%, 71–73% and 83–86% of observations under the three models; part of the network share is structural. Exact rational-arithmetic certificates prove the stage decomposition is unique for all 280 principal network solutions. Regional benchmarking lowers the weaker-commercialization share from 85.7% to 55.4%, a change inseparable from the loss of discrimination of small peer sets.

Keywords: two-stage network DEA; slacks-based measure; linear programming; alternative optima; specification sensitivity; university technology transfer

MSC: 90C08

1. Introduction

1.1. Efficiency Decomposition in Linked Two-Stage Production

Many production systems evaluated with data envelopment analysis (DEA) have an internal structure: an upstream stage generates an intermediate product that a downstream stage consumes. Network DEA represents this structure by linking the stages through the intermediate flow, so a single optimization problem returns system efficiency along with stage efficiencies [1–3]. Comparing the stage efficiencies then identifies the weaker stage, that is, the stage at which measured inefficiency is concentrated [4,5]. However, such a diagnosis is a property of an optimization problem defined on an estimated technology. It is not read directly from the data, and it inherits every modeling choice that defines that problem.

This paper examines that dependence on a concrete two-stage system: provincial university technology transfer in China. Research resources are converted into granted

Check for updates

Academic Editor: Jonathan Blackledge

Received: 18 October 2026
Revised: 3 November 2026
Accepted: 4 November 2026
Published: 6 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

invention patents in a technology-formation stage (Stage F), and these patents are converted into transfer contracts and realized transfer income in a commercialization stage (Stage C) [6,7]. The system is a useful test case for three reasons. Previous studies repeatedly locate inefficiency in the downstream stage [8,9]. Provincial efficiency scores are also used as comparative benchmarks [10]. In addition, the available data are annual provincial aggregates in which neither the production technology nor the gestation lag between the stages is observed directly [11]. Three classes of modeling choices intervene between the data and a stage diagnosis.

1.2. Three Sources of Specification Dependence

First, radial, non-radial and linked representations define different efficiency programs on the same data. The radial Banker–Charnes–Cooper (BCC) model requires all outputs of a stage to expand in the same proportion [12]. The slacks-based measure (SBM) allows each output to adjust by a different proportion [13]. The network SBM additionally requires the two stages to share a common benchmark link flow [3]. Most applications select one representation and interpret efficiency on that basis [7,14]. Few studies, however, separate each step's effect on stage scores, rankings, and classifications using a single fixed sample with matched observation windows.

Second, temporal alignment is a specification rather than an observed lag. Research spending precedes patent grants, which in turn precede contracts and receipts, but annual provincial data do not reveal the gestation period, and a uniform lag across units may be restrictive [15]. Whether the conclusions hold after aligning inputs, intermediate outputs and final outputs contemporaneously and sequentially remains an open measurement question.

Third, the optimal system value is uniquely determined, but alternative optimal solutions can yield different pairs of stage efficiencies with the same value [3,16–18]. A classification read from a solver's decomposition may therefore appear more determinate than the optimization problem justifies. Separately, narrowing or widening the set of peer units changes both the frontier and its power to discriminate among units [19,20].

1.3. Aims, Contributions and Organization

The study addresses two questions. The first is methodological: whether the stage classification produced by the output-oriented network SBM is determinate over the whole set of system-optimal solutions, and which features of the stage scores are fixed by the models themselves. The second is empirical: how scores, rankings and stage classifications respond, on fixed data, to three specification choices, namely, the production representation (radial, non-radial or linked), the temporal alignment of research inputs, patents and commercialization outcomes, and the reference set.

The paper makes four contributions. (i) For the output-oriented two-stage network SBM in envelopment form, we adapt the bounding device that holds system efficiency fixed while maximizing or minimizing the efficiency of one stage, which is established for multiplier-based two-stage models and for a dual decomposition of a two-stage slacks-based model [16,21,22]. Because the network SBM aggregates the stage expansions linearly, the commercialization expansion is an affine function of the formation expansion on the system-optimal face, and the stage gap is strictly monotone in it (Proposition 1). Two auxiliary linear programs per unit therefore bound both stage efficiencies and their gap exactly, and a four-way classification can be verified over all optimal solutions without prioritizing either stage. Non-uniqueness requires the benchmark link flow to vary over the optimal face, and a three-unit example shows that it can. (ii) Ordering properties fix the signs of several specification effects (Remark 3), and the network Stage C efficiency is

measured at the benchmark link flow (Remark 4). (iii) On one fixed sample, we separate the radial-to-non-radial and independent-to-linked steps of the production representation for the alignments $S00$, $S10$ and $S11$, examine the network SBM over a complete 3×3 grid of patent and commercialization lags, and prove the stage decomposition unique for all 280 principal network solutions with exact rational-arithmetic certificates. (iv) The predominant stage classification is stable across the three production specifications under the principal alignments and across the lag grid, whereas its frequency, the scores, the rankings and the frontier projections depend on the specification.

Section 2 reviews related work, Section 3 presents the models and their properties, Section 4 describes the data and the computational design, Section 5 reports the results, Section 6 discusses them, and Section 7 concludes.

2. Related Work

2.1. Radial and Non-Radial Efficiency Measures

DEA evaluates multiple-input, multiple-output production without an explicit functional form. The radial BCC model, which assumes variable returns to scale (VRS), became the standard for benchmarking units against a best-practice frontier [12,23]. SBM models replace the common proportional adjustment with variable-specific adjustments, which matters when outputs do not fall short in the same proportion [13,18,24]. Applied comparisons on Chinese university data have used both families [25–27]. However, the two measures need not differ. When a stage has a single output, the output-oriented radial and slacks-based programs coincide (Remark 1), so comparing them is informative only for multi-output stages.

2.2. Network DEA and the Non-Uniqueness of Divisional Efficiencies

Conventional black-box models do not represent the internal processes that connect inputs to outputs. Network DEA addresses this by incorporating intermediate products and internal links [1–3]. A network model returns a system efficiency as the optimal value of a single program, whereas stage scores are evaluated at an optimal solution, and when the optimal solution is not unique, the stage scores need not be unique either [3,17]. For multiplier-based two-stage models, this ambiguity has been examined by holding system efficiency at its optimum and maximizing or minimizing the efficiency of one stage, under both multiplicative decompositions and additive decompositions in which system efficiency is a weighted average of stage efficiencies [16,21]. An alternative approach removes the ambiguity by imposing a leader–follower priority between the stages [17]. Divisional efficiencies and frontier projections from envelopment and multiplier forms of network models need not coincide, adding another source of interpretive error [28]. For slacks-based network models, the relationship between system and divisional efficiencies has also been analyzed [18,29]. The closest existing approach treats frontier projection and efficiency decomposition in two-stage processes with slacks-based measures; for a dual decomposition, it obtains stage-efficiency intervals by holding the system optimum fixed and maximizing and minimizing one stage efficiency, using an affine relation between the stage terms [22]. The present study applies the same principle to the envelopment form of the output-oriented network SBM with an as-output link, where stage efficiencies are reciprocals of stage expansions, treats the two stages symmetrically and imposes no priority between them. Alternative optima are also a recognized source of ambiguity in DEA-based evaluation more generally [30].

2.3. Temporal Alignment as a Maintained Specification

Dynamic DEA incorporates carry-over activities between successive periods [31]. Dynamic network DEA combines internal links with intertemporal carry-overs [32,33]. Applications show that the temporal treatment affects measured performance, including in university technology transfer, where carry-over efficiency has been assessed jointly with stage and link efficiency [14,34]. A single lag is unlikely to apply to every unit, since production lead times differ [15]. For annual provincial data, moreover, the gestation period connecting research investment, patent grants and transfer receipts is not directly observable [11]. The present study does not estimate a dynamic model; it realigns annual observations and estimates a separate frontier for each starting year, treating the alignment as a maintained specification.

2.4. Specification Sensitivity and Reference-Set Dependence

DEA estimates are relative to an observed production possibility set, and their stability depends on the units evaluated, the indicators chosen and the modeling assumptions adopted [35–37]. In small samples, limited discrimination among frontier units further affects stage diagnosis [20]. Comparative work on efficiency-based benchmarking shows that results are sensitive to the composition of the reference set [38–40]. Observed efficiency differences may also coexist with technological heterogeneity among groups of units [34].

2.5. Application Context: University Technology Transfer as a Two-Stage Process

The innovation-value-chain perspective distinguishes knowledge creation, applied research and commercialization as separate stages through which technological value is generated [41,42]. Evidence from Chinese universities supports this separation: high research and development (R&D) efficiency can coexist with weak transformation outcomes, and the transformation stage is a principal source of overall inefficiency [8,9]. Meta-frontier evidence further shows that technology transfer underperforms R&D across heterogeneous groups of research universities [43]. Two-stage and network models of Chinese university systems have adopted shared-input, dynamic and multi-stage structures [5,7,14,44,45]. Transfer performance is also known to vary with organizational arrangements, transfer channels and regional conditions [4,46–52]. Table 1 summarizes how these strands bear on the present design.

Table 1. Representative literature, its role in the present design and the open measurement question.

Literature Stream	What Is Established	Role in the Present Design	Open Measurement Question
Innovation value chain and Chinese university transformation	Upstream research can perform well while downstream transformation lags [8,41].	Supports a staged design with an explicit weaker stage diagnosis.	Whether stage conclusions are stable across alternative specifications.
Network production structure	Intermediate links can be modeled jointly with system and divisional efficiency [2,3].	Provides the basis for linked-stage evaluation.	Whether linked and independently estimated stages lead to different conclusions.
Non-radial measurement	slacks-based efficiency differs from proportional radial adjustment [13,18].	Motivates comparing BCC, SBM and network SBM on one sample.	Magnitude of specification sensitivity in each stage.
Intertemporal structure and variable lags	Carry-overs matter, and uniform lags can be restrictive [15,31].	Treats timing as a maintained specification, not a known lag.	Whether alternative alignments change scores, ranks and weaker stage classifications.
Specification sensitivity and alternative optima	Scores, rankings and stage decompositions need not be unique [16,18].	Motivates the optimal-face bounds used here.	How far weaker stage classifications depend on the selected decomposition.
Reference-set dependence and regional heterogeneity	Transfer performance differs across regions and reference sets [10,19].	Motivates the regional reference-set comparison.	Whether stage classifications hold when the peer set changes.

3. Models

3.1. Notation and the Two-Stage Production Structure

Tables 2 and 3 define the notation. Each decision-making unit (DMU) is a province in a given starting year. Research resources (H_o, R_o) are converted into patented technology Z_o in Stage F, and Z_o is converted into contracts and income (N_o, I_o) in Stage C. Figure 1a shows this structure together with the link constraint and the stage measures of the network model; Figure 1b summarizes the computational design, that is, the specifications compared, the optimal-face diagnosis computed for each of them, and the quantities reported.

Table 2. Notation: units and observed variables.

Symbol	Meaning
t	Starting year of an observation
o	The decision-making unit being evaluated
j	A reference unit within the frontier
n	Number of decision-making units in one annual frontier
k	Stage index, F for technology formation or C for commercialization
H_o	Full-time-equivalent research personnel, in person-years
R_o	Research and development project expenditure, in thousand RMB at 2018 prices
Z_o	Granted invention patents, in items
Z_j	The same quantity for a reference unit j
H, R, Z, N, I	Row vectors in $\mathbb{R}^{1\times n}_{++}$ collecting the observed values of the n reference units, for example $Z = (Z_1, \dots, Z_n)$
$\mathbf{e}_o$	The o-th standard basis vector of $\mathbb{R}^n$
N_o	Technology-transfer contracts, in number of contracts
I_o	Realized technology-transfer income, in thousand RMB at 2018 prices
A_o	Contract value, in thousand RMB at 2018 prices
$\mathbf{x}_o$	Input vector of the evaluated unit
$\mathbf{y}_o$	Output vector of the evaluated unit
m	Number of inputs
p	Number of outputs
X	Reference input matrix in $\mathbb{R}^{m\times n}_{++}$
Y	Reference output matrix in $\mathbb{R}^{p\times n}_{++}$
$s^-_{H,o}$	Excess in research personnel
$s^-_{R,o}$	Excess in research and development expenditure
$s^+_{Z,o}$	Shortfall in granted invention patents
$s^+_{N,o}$	Shortfall in technology-transfer contracts
$s^+_{I,o}$	Shortfall in realized transfer income
$\mathbf{s}^-$	Vector of input excesses
$\mathbf{s}^+$	Vector of output shortfalls

The two-stage representation requires a single variable that is produced by Stage F and consumed by Stage C. This study uses granted invention patents as that variable for three reasons. First, transactional relevance: a granted invention patent confers an exclusive right that can be licensed, assigned, or contributed as capital, so it is the technology asset that appears in technology-transfer contracts; journal articles and other research outputs are not the object of those transactions and are connected to realized transfer income only indirectly. Second, measurement consistency: granted invention patents are reported annually for each provincial higher-education sector in the same statistical compilation that provides H_o, R_o, N_o and I_o, on a cardinal scale, with a uniform definition and no price dimension [53]. Monetary alternatives are not suitable as the intermediate variable, because contract value and realized income measure the outcomes of commercialization itself; using either as the link would turn part of the commercialization result into the intermediate product and change the production chain that the two stages represent. Third, quality screening

and its cost: applications are filed before substantive examination and include a large share that is never granted, whereas grants have passed examination and are closer to a transferable asset. The cost of this choice is a timing gap: the underlying research precedes filing, and examination takes about a year; this is why temporal alignment is treated as a maintained specification and why the lag grid varies the patent lag over one and two years.

Table 3. Notation: measures, classification and design parameters.

Symbol	Meaning
$\boldsymbol{\lambda}$	Weights assigned to the units of the reference set
$\boldsymbol{\lambda}_F$	The same weights for Stage F
$\boldsymbol{\lambda}_C$	The same weights for Stage C
w_F	Weight of the Stage F
w_C	Weight of the Stage C
ϕ_o	Common proportional expansion of the outputs under the radial benchmark
ϕ_o^*	Optimal proportional expansion under the radial benchmark
q_o	System expansion measure combining the two stages
q_o^*	Optimal system expansion measure
B_o	Output-expansion measure of the slacks-based program, never below one
$B_{F,o}$	Output-expansion measure of Stage F
$B_{C,o}$	Output-expansion measure of Stage C
E_o^{BCC}	Efficiency under the radial benchmark
E_o^{SBM}	Efficiency under the independent slacks-based program
E_o^{NET}	System efficiency under the linked two-stage network program
$E_{o,F}^{\mathrm{NET}}$	Stage F efficiency under that program
$E_{o,C}^{\mathrm{NET}}$	Stage C efficiency under that program
$E_{o,F}^{\min}$, $E_{o,F}^{\max}$	Range of Stage F efficiency compatible with the system optimum
$E_{o,C}^{\min}$, $E_{o,C}^{\max}$	Range of Stage C efficiency compatible with the system optimum
G_o	Difference between the two stage efficiencies
$G_o^{\min}$, $G_o^{\max}$	Range of that difference over the optimal set
$\mathcal{P}_o$	Feasible set of the linear program
$\mathcal{F}_o$	Feasible set of the network program
O_o	System-optimal face: solutions in $\mathcal{F}_o$ attaining q_o^*
$B_{F,o}^{\min}$, $B_{F,o}^{\max}$	Smallest and largest Stage F expansion over O_o
κ_o	Classification label: Stage C weaker, Stage F weaker, equal, or indeterminate
ℓ_Z	Patent lag, in years
ℓ_C	Commercialization lag, in years
L	Length of a rolling patent pool, two or three years
h	Summation index over the years of a rolling pool
a, b	Two specifications being compared
$\mathcal{M}^{a,b}$	Observations available under both specifications
$m_{a,b}$	Number of observations in that matched set
$\mathcal{J}_o$	Reference set of a conditional commercialization program
$\hat{I}_o$	Projected income $I_o + s_{I,o}^+$ under a given program
$\hat{I}_o^{\min}$, $\hat{I}_o^{\max}$	Smallest and largest projected income over the optimal set of that program
δ	Relative relaxation of the optimal-face equality in the numerical check
ε	Absolute tolerance on efficiency differences in the classification rule, 10^{-7}
ρ_s	Spearman rank correlation coefficient
E_o^a, E_o^b	Efficiency (Stage F, Stage C or system) of observation o under specifications a and b
$\Delta_o^{a,b}$	Signed difference $E_o^b - E_o^a$
$\mathrm{MAD}^{a,b}$	Mean absolute difference (MAD) between two specifications
$\Gamma_{o,k}$	Ratio of the stage-k efficiency against the national frontier to that against the regional frontier
$r_{Z,o}$, $r_{N,o}$, $r_{I,o}$	Ratios of the benchmark to the observed patents, contracts and income

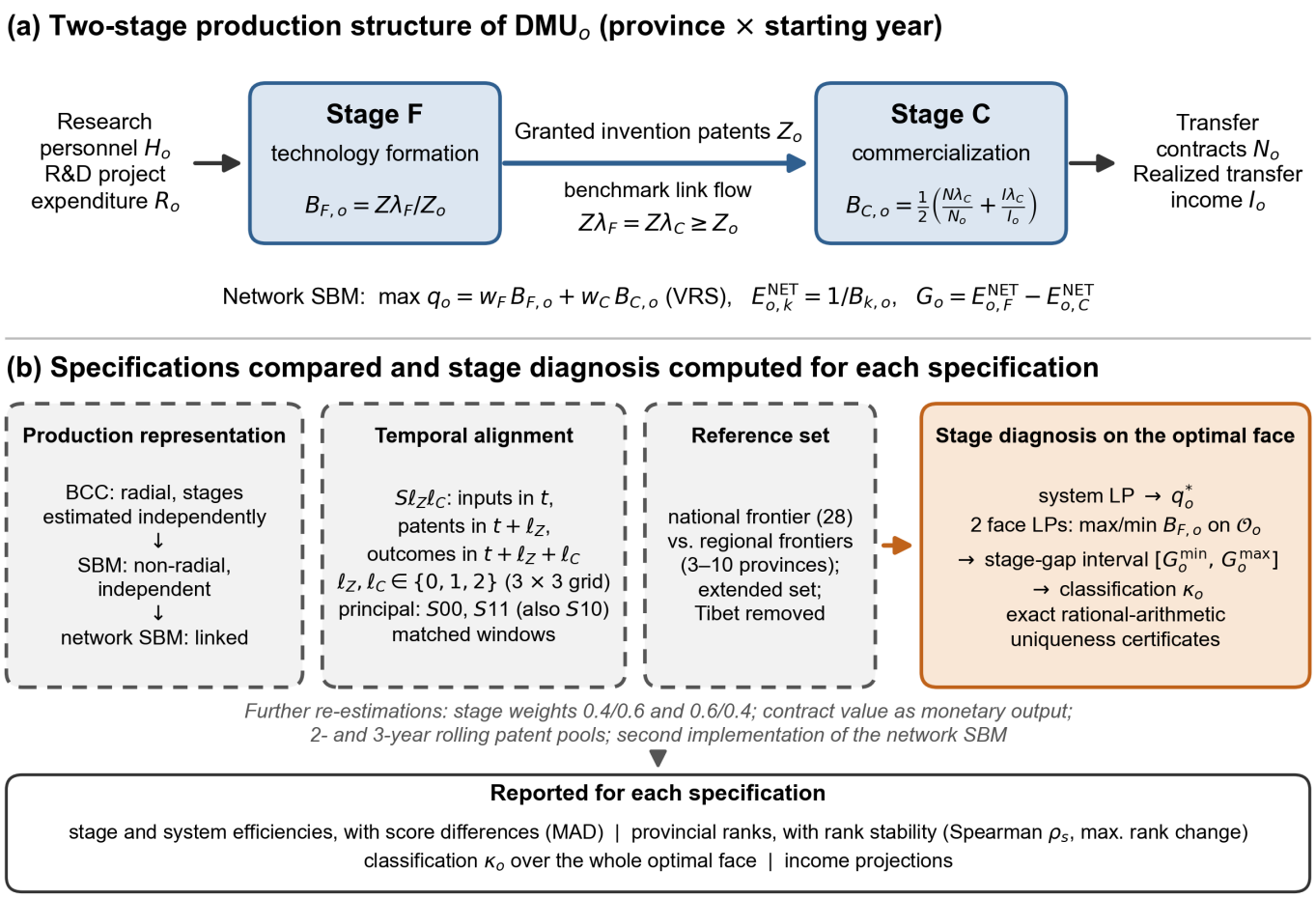


Figure 1. Analytical framework: (**a**) two-stage production structure with the as-output patent link; (**b**) specifications compared, the optimal-face stage diagnosis computed for each specification, and the reported quantities.

The choice also has limitations. A patent is neither necessary nor sufficient for transfer: many granted patents are never licensed, and some transfer occurs through know-how, consulting, or equity in spin-offs that is not patented. The intermediate variable is therefore an observable but incomplete proxy for transferable technological output; the as-output link treatment introduced below assigns any shortfall between the observed patent count and the benchmark link flow to Stage F.

3.2. Output-Oriented VRS BCC Model

We use the output-oriented VRS BCC model as the radial benchmark [12]. The output orientation evaluates how far technological and commercialization outputs can be expanded given the observed resource inputs.

Consider n DMUs. For DMU$_o$, let $\mathbf{x}_o \in \mathbb{R}^m_{++}$ and $\mathbf{y}_o \in \mathbb{R}^p_{++}$ denote its strictly positive observed input and output vectors, let $X \in \mathbb{R}^{m\times n}_{++}$ and $Y \in \mathbb{R}^{p\times n}_{++}$ denote the corresponding reference matrices, which contain the evaluated unit as their o-th column, and let $\boldsymbol{\lambda}$ denote the intensity vector. Strict positivity keeps every ratio in the slacks-based objectives below well defined. The output-oriented VRS BCC model is

$$\begin{aligned} \max_{\phi_o, \boldsymbol{\lambda}} \quad & \phi_o \\ \text{s.t.} \quad & X\boldsymbol{\lambda} \le \mathbf{x}_o, \\ & Y\boldsymbol{\lambda} \ge \phi_o \mathbf{y}_o, \\ & \mathbf{1}^\top \boldsymbol{\lambda} = 1, \\ & \boldsymbol{\lambda} \ge \mathbf{0}, \end{aligned} \tag{1}$$

and the corresponding efficiency score is

$$E_o^{\mathrm{BCC}} = \frac{1}{\phi_o^*}, \qquad 0 < E_o^{\mathrm{BCC}} \le 1, \tag{2}$$

where ϕ_o^* denotes the optimal proportional output-expansion factor; $\phi_o^* \ge 1$ because $\boldsymbol{\lambda} = \mathbf{e}_o$, $\phi_o = 1$ is feasible.

For Stage F, $\mathbf{x}_o = (H_o, R_o)^\top$ and $y_o = Z_o$. For Stage C, $x_o = Z_o$ and $\mathbf{y}_o = (N_o, I_o)^\top$. The Stage C BCC model therefore imposes a common proportional expansion factor on the two commercialization outputs. The two stages are estimated independently, and no restriction requires the Stage F and Stage C benchmark values of Z_o to coincide.

3.3. Independent Output-Oriented SBM

The radial BCC specification requires all outputs to expand proportionally, which can be restrictive when a stage has several outputs with heterogeneous shortfalls. The second specification therefore adopts an output-oriented SBM [13]. For a stage with p outputs, let $\mathbf{s}^-$ and $\mathbf{s}^+$ denote input excesses and output shortfalls, respectively. The independent output-oriented VRS SBM is

$$\begin{aligned} \max_{\boldsymbol{\lambda}, \mathbf{s}^-, \mathbf{s}^+} \quad & B_o = 1 + \frac{1}{p} \sum_{r=1}^{p} \frac{s_{ro}^+}{y_{ro}} \\ \text{s.t.} \quad & \mathbf{x}_o = X\boldsymbol{\lambda} + \mathbf{s}^-, \\ & \mathbf{y}_o = Y\boldsymbol{\lambda} - \mathbf{s}^+, \\ & \mathbf{1}^\top \boldsymbol{\lambda} = 1, \\ & \boldsymbol{\lambda}, \mathbf{s}^-, \mathbf{s}^+ \ge \mathbf{0}, \end{aligned} \tag{3}$$

with efficiency score

$$E_o^{\mathrm{SBM}} = \frac{1}{B_o^*}. \tag{4}$$

Because the objective is linear in the output slacks and all constraints are linear, Equation (3) is a linear program; no fractional transformation is required in the output orientation. Input slacks enter the feasibility conditions but not the objective, and we again estimate the two stages independently.

Remark 1 (Single-output stages)**.** *If a stage has a single output ($p = 1$), the output constraint of Equation (3) gives $s_o^+ = Y\boldsymbol{\lambda} - y_o$, so $B_o = Y\boldsymbol{\lambda}/y_o$. Both Equation (1) and Equation (3) then maximize $Y\boldsymbol{\lambda}/y_o$ over $\{\boldsymbol{\lambda} \ge \mathbf{0} : X\boldsymbol{\lambda} \le \mathbf{x}_o,\ \mathbf{1}^\top\boldsymbol{\lambda} = 1\}$; the additional requirement $Y\boldsymbol{\lambda} \ge y_o$ in the SBM, which arises from $s_o^+ \ge 0$, is redundant for the maximization (removing it changes neither the optimal value nor the optimal solution set), because $\boldsymbol{\lambda} = \mathbf{e}_o$ is feasible and attains the value one. Hence, $B_o^* = \phi_o^*$ and $E_o^{\mathrm{SBM}} = E_o^{\mathrm{BCC}}$.*

Because Stage F has the single output Z_o, its BCC and SBM scores coincide by construction, and the BCC–SBM comparison is informative only for Stage C, where the SBM allows N_o and I_o to exhibit different proportional shortfalls.

3.4. Two-Stage Network SBM with an As-Output Link

The independent BCC and SBM specifications do not require the benchmark patent output of Stage F to equal the benchmark patent input of Stage C. To incorporate this internal structure, the third specification adopts an output-oriented two-stage network SBM [3]. Let $\boldsymbol{\lambda}_F$ and $\boldsymbol{\lambda}_C$ denote the stage-specific intensity vectors, and let $H, R, Z, N, I \in \mathbb{R}_{++}^{1\times n}$ denote the row vectors of observed values across the n reference units. The network production technology for DMU_o is defined by

$$\begin{gathered} H_o = H\boldsymbol{\lambda}_F + s_{H,o}^-, \qquad R_o = R\boldsymbol{\lambda}_F + s_{R,o}^-, \\ Z_o = Z\boldsymbol{\lambda}_F - s_{Z,o}^+, \qquad N_o = N\boldsymbol{\lambda}_C - s_{N,o}^+, \qquad I_o = I\boldsymbol{\lambda}_C - s_{I,o}^+, \\ Z\boldsymbol{\lambda}_F = Z\boldsymbol{\lambda}_C, \\ \mathbf{1}^\top\boldsymbol{\lambda}_F = 1, \qquad \mathbf{1}^\top\boldsymbol{\lambda}_C = 1, \\ \boldsymbol{\lambda}_F, \boldsymbol{\lambda}_C, s_{H,o}^-, s_{R,o}^-, s_{Z,o}^+, s_{N,o}^+, s_{I,o}^+ \geq 0. \end{gathered} \tag{5}$$

The equality $Z\boldsymbol{\lambda}_F = Z\boldsymbol{\lambda}_C$ is the intermediate-link continuity condition. It requires the benchmark quantity of technological output generated by Stage F to equal the quantity entering Stage C. We call this common quantity, $Z\boldsymbol{\lambda}_F = Z\boldsymbol{\lambda}_C$, the benchmark link flow; the two stages therefore cannot select mutually inconsistent benchmark link flows. The intermediate variable Z_o is treated as an as-output link: it is an internal network flow, not an ordinary external Stage C input with an additional input slack. The shortfall between the observed intermediate output and its benchmark quantity enters the objective only through the supplying stage, Stage F. It nevertheless affects Stage C because the link-continuity constraint imposes the same benchmark link flow on Stage C (Remark 4).

The stage-specific output-expansion measures are

$$B_{F,o} = 1 + \frac{s_{Z,o}^+}{Z_o}, \qquad B_{C,o} = 1 + \frac{1}{2}\left(\frac{s_{N,o}^+}{N_o} + \frac{s_{I,o}^+}{I_o}\right). \tag{6}$$

Let w_F and w_C denote pre-specified positive stage weights satisfying $w_F + w_C = 1$. Conditional on these weights, the system-level network SBM solves

$$\max_{\boldsymbol{\lambda}_F, \boldsymbol{\lambda}_C, \mathbf{s}_o} \quad q_o = w_F B_{F,o} + w_C B_{C,o} \tag{7}$$

subject to all production, VRS, non-negativity and link-continuity constraints in Equation (5), where the decision vector is $\xi = (\boldsymbol{\lambda}_F, \boldsymbol{\lambda}_C, \mathbf{s}_o)$ with composite slack vector $\mathbf{s}_o = (s_{H,o}^-, s_{R,o}^-, s_{Z,o}^+, s_{N,o}^+, s_{I,o}^+)^\top \in \mathbb{R}_+^5$. We write $\mathcal{F}_o$ for the feasible set defined by these constraints and q_o^* for the optimal value. Evaluated at an optimal solution ξ of Equation (7), the system and stage efficiencies are

$$E_o^{\mathrm{NET}} = \frac{1}{q_o^*}, \qquad E_{o,F}^{\mathrm{NET}} = \frac{1}{B_{F,o}}, \qquad E_{o,C}^{\mathrm{NET}} = \frac{1}{B_{C,o}}, \tag{8}$$

and they satisfy

$$\frac{1}{E_o^{\mathrm{NET}}} = \frac{w_F}{E_{o,F}^{\mathrm{NET}}} + \frac{w_C}{E_{o,C}^{\mathrm{NET}}}. \tag{9}$$

Remark 2 (Linear form and harmonic aggregation)**.** *(a) Eliminating the slacks through the equalities in Equation (5) gives* $B_{F,o} = Z\boldsymbol{\lambda}_F/Z_o$ *and* $B_{C,o} = \frac{1}{2}(N\boldsymbol{\lambda}_C/N_o + I\boldsymbol{\lambda}_C/I_o)$*, so Equation (7) is the linear program*

$$\begin{aligned} \max_{\boldsymbol{\lambda}_F, \boldsymbol{\lambda}_C \geq \mathbf{0}} \quad & w_F \sum_j \lambda_{F,j}\frac{Z_j}{Z_o} + \frac{w_C}{2}\sum_j \lambda_{C,j}\left(\frac{N_j}{N_o} + \frac{I_j}{I_o}\right) \\ s.t. \quad & H\boldsymbol{\lambda}_F \leq H_o, \quad R\boldsymbol{\lambda}_F \leq R_o, \quad Z\boldsymbol{\lambda}_F \geq Z_o, \\ & N\boldsymbol{\lambda}_C \geq N_o, \quad I\boldsymbol{\lambda}_C \geq I_o, \quad Z\boldsymbol{\lambda}_F = Z\boldsymbol{\lambda}_C, \\ & \mathbf{1}^\top\boldsymbol{\lambda}_F = 1, \quad \mathbf{1}^\top\boldsymbol{\lambda}_C = 1. \end{aligned} \tag{10}$$

By Lemma 1 below, Equations (7) and (10) have the same optimal value and their optimal sets correspond one to one; $\mathcal{F}_o$ is a nonempty compact polyhedron, so q_o^* is attained, and $q_o^* \geq 1$ because the solution with $\boldsymbol{\lambda}_F = \boldsymbol{\lambda}_C = \mathbf{e}_o$ is feasible with objective value one. (b) Equation (9) states that E_o^{NET} is the weighted harmonic mean of the stage efficiencies with weights (w_F, w_C). Hence,

$$\min\{E_{o,F}^{\mathrm{NET}}, E_{o,C}^{\mathrm{NET}}\} \leq E_o^{\mathrm{NET}} \leq w_F E_{o,F}^{\mathrm{NET}} + w_C E_{o,C}^{\mathrm{NET}}, \tag{11}$$

with equality throughout if and only if $E_{o,F}^{\mathrm{NET}} = E_{o,C}^{\mathrm{NET}}$; and since $B_{F,o}, B_{C,o} \geq 1$, $E_o^{\mathrm{NET}} = 1$ if and only if both stage efficiencies equal one.

Lemma 1 (Equivalence of the slack and intensity formulations)**.** *Let $\mathcal{P}_o$ denote the feasible set of Equation (10), and define $\Phi(\boldsymbol{\lambda}_F, \boldsymbol{\lambda}_C) = (\boldsymbol{\lambda}_F, \boldsymbol{\lambda}_C, \mathbf{s}_o(\boldsymbol{\lambda}_F, \boldsymbol{\lambda}_C))$ with $s_{H,o}^- = H_o - H\boldsymbol{\lambda}_F$, $s_{R,o}^- = R_o - R\boldsymbol{\lambda}_F$, $s_{Z,o}^+ = Z\boldsymbol{\lambda}_F - Z_o$, $s_{N,o}^+ = N\boldsymbol{\lambda}_C - N_o$ and $s_{I,o}^+ = I\boldsymbol{\lambda}_C - I_o$. Then: (i) Φ is an affine bijection from $\mathcal{P}_o$ onto $\mathcal{F}_o$, whose inverse is the coordinate projection $\pi(\xi) = (\boldsymbol{\lambda}_F, \boldsymbol{\lambda}_C)$; (ii) $q_o(\Phi(\boldsymbol{\lambda}_F, \boldsymbol{\lambda}_C))$ equals the objective of Equation (10) at $(\boldsymbol{\lambda}_F, \boldsymbol{\lambda}_C)$, so the two programs have the same optimal value, Φ maps the optimal set of Equation (10) onto that of Equation (7), and Φ maps faces and extreme points of $\mathcal{P}_o$ onto faces and extreme points of $\mathcal{F}_o$; (iii) $\mathcal{P}_o$ and $\mathcal{F}_o$ are nonempty compact polyhedra.*

Proof. (i) Each component of $\mathbf{s}_o$ is affine in $(\boldsymbol{\lambda}_F, \boldsymbol{\lambda}_C)$, so Φ is affine. Let $(\boldsymbol{\lambda}_F, \boldsymbol{\lambda}_C) \in \mathcal{P}_o$. The five inequalities $H\boldsymbol{\lambda}_F \leq H_o$, $R\boldsymbol{\lambda}_F \leq R_o$, $Z\boldsymbol{\lambda}_F \geq Z_o$, $N\boldsymbol{\lambda}_C \geq N_o$ and $I\boldsymbol{\lambda}_C \geq I_o$ of Equation (10) are exactly $\mathbf{s}_o \geq \mathbf{0}$; the five production equalities of Equation (5) hold by the definition of $\mathbf{s}_o$; and the link, convexity and non-negativity constraints on $(\boldsymbol{\lambda}_F, \boldsymbol{\lambda}_C)$ are common to both formulations. Hence, $\Phi(\boldsymbol{\lambda}_F, \boldsymbol{\lambda}_C) \in \mathcal{F}_o$. Conversely, let $\xi = (\boldsymbol{\lambda}_F, \boldsymbol{\lambda}_C, \mathbf{s}_o) \in \mathcal{F}_o$. Each production equality of Equation (5) contains exactly one slack, with coefficient ± 1, and therefore determines that slack uniquely; hence, $\mathbf{s}_o = \mathbf{s}_o(\boldsymbol{\lambda}_F, \boldsymbol{\lambda}_C)$ and $\xi = \Phi(\pi(\xi))$. Because $\mathbf{s}_o \geq \mathbf{0}$, $\pi(\xi)$ satisfies the five inequalities of Equation (10), so $\pi(\xi) \in \mathcal{P}_o$. Therefore, $\Phi(\mathcal{P}_o) = \mathcal{F}_o$, $\pi \circ \Phi$ is the identity on $\mathcal{P}_o$ and $\Phi \circ \pi$ is the identity on $\mathcal{F}_o$, so Φ is a bijection with linear inverse π. (ii) Substituting $\mathbf{s}_o$ into Equation (6) gives $B_{F,o} = 1 + (Z\boldsymbol{\lambda}_F - Z_o)/Z_o = Z\boldsymbol{\lambda}_F/Z_o$ and $B_{C,o} = \frac{1}{2}(N\boldsymbol{\lambda}_C/N_o + I\boldsymbol{\lambda}_C/I_o)$, which yields the objective of Equation (10). A bijection that preserves objective values preserves optimal values and maps optimal sets onto optimal sets. An affine bijection preserves convex combinations and therefore maps faces onto faces and extreme points onto extreme points; in particular, the face programs in Equation (14) may be solved over $\mathcal{P}_o$. (iii) $\mathcal{P}_o$ is defined by finitely many linear constraints and lies in the product of two unit simplices, so it is a compact polyhedron; it is nonempty because $(\mathbf{e}_o, \mathbf{e}_o)$ satisfies every constraint with all slacks equal to zero. $\mathcal{F}_o = \Phi(\mathcal{P}_o)$ is the affine image of a polytope and is therefore a polytope. □

Remark 2(a) gives the program that is solved; part (b) bounds the system efficiency of each unit below by the smaller stage efficiency and above by the weighted arithmetic mean of the two. The three production specifications thus differ in distinct ways. The BCC-to-SBM comparison shifts output adjustment from radial to non-radial while retaining independent-stage estimation. The SBM-to-network-SBM comparison captures the joint effect of system-level optimization and of a common benchmark link flow, which may exceed the observed patent count. A direct BCC-to-network-SBM comparison changes both features at once. Two further properties of the three programs fix the direction of several of these differences.

Remark 3 (Ordering of stage scores across specifications)**.** *(a) For either stage, an optimal $\boldsymbol{\lambda}^*$ of Equation (1) satisfies $X\boldsymbol{\lambda}^* \leq \mathbf{x}_o$ and $Y\boldsymbol{\lambda}^* \geq \phi_o^* \mathbf{y}_o \geq \mathbf{y}_o$, so it is feasible in Equation (3) with $\mathbf{s}^+ = Y\boldsymbol{\lambda}^* - \mathbf{y}_o$, and every output ratio $(Y\boldsymbol{\lambda}^*)_r / y_{ro}$ is at least ϕ_o^*. Hence, $B_o^* \geq \phi_o^*$ and $E_o^{\mathrm{SBM}} \leq E_o^{\mathrm{BCC}}$. (b) The $\boldsymbol{\lambda}_F$ component of every $\xi \in \mathcal{F}_o$ satisfies $H\boldsymbol{\lambda}_F \leq H_o$, $R\boldsymbol{\lambda}_F \leq R_o$, $Z\boldsymbol{\lambda}_F \geq Z_o$ and $\mathbf{1}^\top \boldsymbol{\lambda}_F = 1$,*

and is therefore feasible in the independent Stage F program. Hence, $B_{F,o}(\xi)$ cannot exceed the optimal value B_o^* of the independent Stage F program, that is, $E_{o,F}^{\mathrm{NET}}(\xi) \geq E_{o,F}^{\mathrm{SBM}} = E_{o,F}^{\mathrm{BCC}}$ at every feasible and in particular every optimal solution. (c) No ordering holds in general for Stage C. The independent program admits Stage C benchmarks with $Z\boldsymbol{\lambda} \leq Z_o$, whereas the network program requires $Z\boldsymbol{\lambda}_C = Z\boldsymbol{\lambda}_F \geq Z_o$; the two sets of admissible benchmarks intersect only where $Z\boldsymbol{\lambda} = Z_o$, so changes of either sign are possible.

Remark 4 (Benchmark link flow and the Stage C efficiency)**.** *Since $B_{F,o}(\xi) = Z\boldsymbol{\lambda}_F/Z_o$ on $\mathcal{F}_o$, Stage F is inefficient at ξ exactly when the benchmark link flow exceeds the observed patent count, $Z\boldsymbol{\lambda}_F > Z_o$. Stage C is then benchmarked against reference combinations that use $Z\boldsymbol{\lambda}_C = Z\boldsymbol{\lambda}_F > Z_o$ patents. The network Stage C efficiency is therefore measured at the benchmark link flow, whereas the independent SBM Stage C efficiency is measured at the observed patent count. Two consequences follow. (a) If $Z\boldsymbol{\lambda}_F = Z_o$ at every system-optimal solution, then $E_{o,F}^{\mathrm{NET}} = 1$ and $G_o \geq 0$ on the whole optimal face, so Equation (16) labels the unit "Stage C weaker" or "equal". (b) A "Stage F weaker" label requires $G_o < -\varepsilon$ on the whole optimal face, so the link must be expanded at every system-optimal solution. Example 1 shows that the quantifier in (b) matters: there the link is not expanded at one optimal solution, yet the unit is indeterminate.*

By Remark 3, the signs of the Stage F change from the independent SBM to the network SBM and of the Stage C change from BCC to SBM are determined by the models, and only their magnitudes are empirical; the sign of the Stage C change from the independent SBM to the network SBM is not determined and is an empirical outcome.

3.5. Stage-Efficiency Bounds on the System-Optimal Face

The network SBM returns a system efficiency and two stage efficiencies, and their difference identifies the weaker stage:

$$G_o = E_{o,F}^{\mathrm{NET}} - E_{o,C}^{\mathrm{NET}}. \tag{12}$$

The optimal value q_o^* of Equation (7) is uniquely determined, but the set of solutions attaining it may contain more than one point, and different optimal solutions may yield different stage efficiencies. A classification based on the single solution a solver returns may therefore depend on which optimal solution it returns. Instead, we evaluate the full range of stage efficiencies compatible with the optimal value. Define the system-optimal face

$$O_o = \left\{\xi \in \mathcal{F}_o : \; w_F B_{F,o}(\xi) + w_C B_{C,o}(\xi) = q_o^*\right\} \tag{13}$$

and the two auxiliary linear programs

$$B_{F,o}^{\max} = \max_{\xi \in O_o} B_{F,o}(\xi), \qquad B_{F,o}^{\min} = \min_{\xi \in O_o} B_{F,o}(\xi). \tag{14}$$

We obtain each by appending the equality in Equation (13) to the constraints of Equation (10) and changing the objective. Since O_o is a nonempty face of the compact polyhedron $\mathcal{F}_o$, both optima are attained.

Proposition 1. *On O_o, $B_{C,o} = (q_o^* - w_F B_{F,o})/w_C$, and the stage gap $G_o = 1/B_{F,o} - 1/B_{C,o}$ is a strictly decreasing function of $B_{F,o}$. Consequently, over O_o,*

$$\begin{gathered} E_{o,F}^{\min} = \frac{1}{B_{F,o}^{\max}}, \qquad E_{o,F}^{\max} = \frac{1}{B_{F,o}^{\min}}, \qquad E_{o,C}^{\min} = \frac{w_C}{q_o^* - w_F B_{F,o}^{\min}}, \qquad E_{o,C}^{\max} = \frac{w_C}{q_o^* - w_F B_{F,o}^{\max}}, \\ G_o^{\min} = E_{o,F}^{\min} - E_{o,C}^{\max}, \qquad G_o^{\max} = E_{o,F}^{\max} - E_{o,C}^{\min}, \end{gathered} \tag{15}$$

and every value between these endpoints is attained.

Proof. The identity for $B_{C,o}$ follows from the defining equality of O_o. Since all slacks in Equation (5) are non-negative, $B_{F,o} \geq 1$ and $B_{C,o} \geq 1$ on $\mathcal{F}_o$. Writing $g(B) = 1/B - w_C/(q_o^* - w_F B)$, we therefore have $q_o^* - w_F B = w_C B_{C,o} \geq w_C > 0$ on O_o and $g'(B) = -1/B^2 - w_F w_C/(q_o^* - w_F B)^2 < 0$. Because O_o is convex and compact and $B_{F,o}$ is linear, $B_{F,o}(O_o) = [B_{F,o}^{\min}, B_{F,o}^{\max}]$. The map $B \mapsto 1/B$ is decreasing and $B \mapsto w_C/(q_o^* - w_F B)$ is increasing, which gives the stage endpoints; strict monotonicity and continuity of g give the endpoints and the full range of G_o. □

Holding the system value at its optimum and optimizing one stage is an established device for multiplier-based two-stage models with multiplicative and additive decompositions [16,21], and it has been applied to the dual decomposition of a two-stage slacks-based model [22]. Proposition 1 is its counterpart for the envelopment form of the slacks-based model with an as-output link. It adds the monotonicity of the stage gap, on which the exact classification of Corollary 1 rests.

The proposition replaces the bounding of the stage gap, a non-linear function of the solution, with two linear programs. One system program and two face programs per unit therefore make the stage diagnosis a verifiable computation. The reduction relies on the affine aggregation of the stage expansions in Equation (7) and does not extend directly to network formulations with other aggregation structures.

The sign of the gap and the width of its interval are read together to classify an evaluated unit as

$$\kappa_o = \begin{cases} \text{Stage C weaker}, & G_o^{\min} > \varepsilon, \\ \text{Stage F weaker}, & G_o^{\max} < -\varepsilon, \\ \text{equal}, & [G_o^{\min}, G_o^{\max}] \subseteq [-\varepsilon, \varepsilon], \\ \text{indeterminate}, & \text{otherwise}, \end{cases} \tag{16}$$

where $\varepsilon \geq 0$ is an absolute tolerance on efficiency differences; with $\varepsilon = 0$ the rule refers to the exact interval.

Corollary 1. *With $G_o^{\min}$ and $G_o^{\max}$ given by Equation (15), a unit is labeled "Stage C weaker" by Equation (16) if and only if $G_o > \varepsilon$ at every system-optimal solution, "Stage F weaker" if and only if $G_o < -\varepsilon$ at every system-optimal solution, and "indeterminate" if and only if neither these conditions nor the equality condition hold over the whole of O_o.*

Proof. By Proposition 1, the set of values taken by G_o on O_o is exactly the interval $[G_o^{\min}, G_o^{\max}]$. Hence, $G_o > \varepsilon$ on all of O_o if and only if $G_o^{\min} > \varepsilon$, $G_o < -\varepsilon$ on all of O_o if and only if $G_o^{\max} < -\varepsilon$, and $|G_o| \leq \varepsilon$ on all of O_o if and only if $[G_o^{\min}, G_o^{\max}] \subseteq [-\varepsilon, \varepsilon]$. The indeterminate case is the complement of these three. □

The bounds are non-trivial only under a specific geometric condition, which may explain why they are often degenerate in applications.

Remark 5 (When is the stage decomposition non-unique?)**.** *(a) Geometric condition. On O_o, $B_{F,o} = Z\boldsymbol{\lambda}_F/Z_o$, so $B_{F,o}^{\min} < B_{F,o}^{\max}$ if and only if the benchmark link flow $Z\boldsymbol{\lambda}_F$ is not constant on O_o. Because O_o is the face of $\mathcal{F}_o$ on which the linear objective attains its maximum, this requires a feasible direction along O_o in which the link flow changes while the objective does not, that is, $w_F \Delta B_{F,o} = -w_C \Delta B_{C,o}$. (b) Genericity. Such a flat trade-off requires an exact equality between a data-determined rate and the weight ratio w_F/w_C. It may therefore be nongeneric under continuously varying data, but it remains possible, as Example 1 shows. (c) Basis criterion. At an optimal basis, the optimal solution, and hence the decomposition, is unique if every nonbasic reduced cost is strictly negative [54]. A zero nonbasic reduced cost (dual degeneracy) is necessary for non-uniqueness but*

not sufficient for a non-unique decomposition, because the alternative optima it admits may leave $Z\boldsymbol{\lambda}_F$ unchanged. If $q_o^* = 1$, the decomposition is trivially unique, because $B_{F,o} = B_{C,o} = 1$ is then forced. (d) Output mix. Uniqueness of the pair $(E_{o,F}^{\mathrm{NET}}, E_{o,C}^{\mathrm{NET}})$ does not imply uniqueness of $\boldsymbol{\lambda}_C$ or of the projected output mix; the income projection is therefore bounded separately in Equation (17).

Example 1. *Let three units have $H_j = R_j = 1$ and (Z_j, N_j, I_j) equal to $(1,1,1)$ for the evaluated unit o, $(2,1,1)$ for unit A and $(1,2,2)$ for unit B, with $w_F = w_C = 1/2$. Every $\boldsymbol{\lambda}_F$ in the unit simplex is feasible for Stage F, and the benchmark link $\bar{Z} = Z\boldsymbol{\lambda}_F = 1 + \lambda_{F,A}$ ranges over $[1,2]$. Given $Z\boldsymbol{\lambda}_C = \bar{Z}$ and $\mathbf{1}^\top\boldsymbol{\lambda}_C = 1$, we have $1 + \lambda_{C,A} = \bar{Z}$, so $\lambda_{C,A} = \bar{Z} - 1$ and $\lambda_{C,B} = 2 - \bar{Z} - \lambda_{C,o} \le 2 - \bar{Z}$. Thus $N\boldsymbol{\lambda}_C = I\boldsymbol{\lambda}_C = 1 + \lambda_{C,B} \le 3 - \bar{Z}$ and $q_o \le \frac{1}{2}\bar{Z} + \frac{1}{2}(3 - \bar{Z}) = 1.5$, with equality for every $\bar{Z} \in [1,2]$ at $\boldsymbol{\lambda}_C = (0, \bar{Z} - 1, 2 - \bar{Z})^\top$, the components ordered as (o, A, B). The optimal face therefore contains decompositions with $B_{F,o} \in [1,2]$ and $B_{C,o} = 3 - B_{F,o}$, so $E_{o,F}^{\mathrm{NET}}, E_{o,C}^{\mathrm{NET}} \in [0.5, 1]$ and $G_o \in [-0.5, 0.5]$, and the unit is classified as indeterminate for any $\varepsilon < 1/2$. The solver-returned decomposition could report either stage as the weaker one. For these data, the face programs in Equation (14) return exactly these intervals.*

The classification imposes no tie-breaking rule among alternative optima and rests on the bounds alone; where single stage scores and projections are reported, they are taken at the optimal solution that maximizes $B_{F,o}$. For the independently estimated BCC and SBM specifications, each stage score is the optimal value of its own program and is therefore unique, so Equation (16) is applied with $G_o^{\min} = G_o^{\max}$. In computation, q_o^* is known only to solver tolerance, and the face programs impose the equality at the computed optimum.

3.6. Projection Bounds and Conditional Commercialization Programs

A unique stage-efficiency pair does not by itself guarantee a unique composition of the two commercialization outputs, because a solution that expands contracts more and income less can attain the same Stage C expansion as one that does the opposite. The income projection $\hat{I}_o = I_o + s_{I,o}^+ = I\boldsymbol{\lambda}_C$ is therefore bounded over the same optimal face:

$$\hat{I}_o^{\min} = \min_{\zeta \in O_o} I\boldsymbol{\lambda}_C, \qquad \hat{I}_o^{\max} = \max_{\zeta \in O_o} I\boldsymbol{\lambda}_C. \tag{17}$$

If $\hat{I}_o^{\max} - \hat{I}_o^{\min}$ is negligible, the reported income shortfall does not depend on which optimal solution the solver returns.

Because the benchmark link flow may exceed the observed patent count, the network projection need not be attainable with the observed annual patent flow. We therefore compute two further variants with nested feasible sets, the second contained in the first; the income projection of an individual unit need not change monotonically between them. Both are standalone Stage C programs that differ only in the reference set they admit. Writing $\mathcal{J}_o$ for that set, each program is

$$\begin{gathered}
\max_{\{\lambda_j\}_{j\in\mathcal{J}_o},\, s_{N,o}^+,\, s_{I,o}^+} \frac{1}{2}\left(\frac{s_{N,o}^+}{N_o} + \frac{s_{I,o}^+}{I_o}\right) \\
\text{subject to} \\
\sum_{j\in\mathcal{J}_o} \lambda_j Z_j \le Z_o, \qquad \sum_{j\in\mathcal{J}_o} \lambda_j N_j - s_{N,o}^+ = N_o, \\
\sum_{j\in\mathcal{J}_o} \lambda_j I_j - s_{I,o}^+ = I_o, \qquad \sum_{j\in\mathcal{J}_o} \lambda_j = 1, \\
\lambda_j,\, s_{N,o}^+,\, s_{I,o}^+ \ge 0,
\end{gathered} \tag{18}$$

and the projected income is $\hat{I}_o = I_o + s_{I,o}^+$. The two output rows imply $N\boldsymbol{\lambda} \ge N_o$ and $I\boldsymbol{\lambda} \ge I_o$, so neither output can fall below its observed level, independently of which optimum the solver returns. The two variants differ only in the reference set $\mathcal{J}_o$:

- the first takes all provinces of the year, which makes it the income projection of the independent Stage C program of Equation (3) once its input slack is written as the bound $Z\boldsymbol{\lambda} \leq Z_o$;
- the second admits only provinces whose observed patent count does not exceed that of the evaluated province, $Z_j \leq Z_o$, so a province is benchmarked against provinces granted no more invention patents in the same year, while always retaining the evaluated unit to keep the program feasible.

The first variant bounds the weighted patent flow of the benchmark, so a province with more patents than the evaluated one may still carry weight; the second restricts each candidate peer. Both restrictions apply to the annual patent flow, since these data do not observe the patent stock. Neither program evaluates technology formation, so neither returns a Stage F efficiency or a stage classification. The first program reproduces the independent SBM Stage C efficiency; the second has a province-specific reference set and is therefore a conservative comparator for the income projection. For each conditional program, the income projection is bounded over its own optimal set as in Equation (17).

3.7. Reference-Set Nesting

We compute each score against a single national frontier built from all twenty-eight provinces. We also recompute the same scores for each province against a frontier built only from provinces in its own region, and compute the ratio of the national to the regional score:

$$\Gamma_{o,k} = \frac{E_{o,k}^{\text{national}}}{E_{o,k}^{\text{regional}}}, \qquad k \in \{F, C\}. \tag{19}$$

Remark 6 (Reference-set nesting)**.** *Let $\mathcal{J} \subseteq \mathcal{J}'$ be two reference sets that both contain the evaluated unit o. Every feasible solution of Equation (10) under $\mathcal{J}$ is feasible under $\mathcal{J}'$ after setting the intensities of the added units to zero, so $q_o^*(\mathcal{J}) \leq q_o^*(\mathcal{J}')$ and $E_o^{\text{NET}}(\mathcal{J}') \leq E_o^{\text{NET}}(\mathcal{J})$. No such monotonicity holds for the individual stage scores, since enlarging the feasible set can raise the weighted objective without raising both stage expansions.*

By Remark 6, the system efficiency cannot rise when the regional reference set is replaced by the larger national one, whereas $\Gamma_{o,k}$ in Equation (19) carries no theoretical bound and is descriptive. A regional frontier estimated from few units is loose, so the companion quantity to read alongside the ratio is the share of observations that lie on the regional frontier. Here, a unit is said to lie on a frontier when $E_o^{\text{NET}} = 1$, equivalently when both stage efficiencies equal one (Remark 2(b)); this is efficiency in the output-oriented sense, which does not price input slacks. When the on-frontier share is high, the regional benchmark has little discriminating power, and changes in classification under it cannot be attributed to technology differences.

4. Computational Design

4.1. Data, Sample and Deflation

The data are provincial-level statistics on science and technology activities of Chinese higher education institutions from 2018 to 2023, taken from the Ministry of Education of China's annual higher-education science and technology statistical compilations [53]. The source panel covers 31 provincial-level regions of mainland China with one observation per province-year, giving 186 province-year records before sample screening. Three western provinces are excluded from the fixed sample: Xinjiang and Qinghai report no technology-transfer contracts and no realized transfer income in 2018, and Tibet reports none in 2018,

2019 and 2023 (Table A3 in Appendix C). Because contracts and income are denominators of the slacks-based ratios, these province-years cannot be evaluated. We do not replace the zeros with small positive constants, which would make the ratios depend on an arbitrary constant and create exactly the extreme small-denominator ratios examined in Section 5.5. All other records of the panel are strictly positive, which leaves a fixed sample of 28 provinces used throughout. The three provinces enter the extended reference set of Section 5.7 in every starting year in which their records are positive.

All monetary quantities are expressed in 2018 prices using the national consumer price index, which is 1.000 in 2018, 1.029 in 2019, 1.055 in 2020, 1.064 in 2021, 1.086 in 2022 and 1.088 in 2023, as published in the annual statistical communiqués of the National Bureau of Statistics of China [55]. Each year's nominal value is divided by its index. The indices are shown rounded to three decimals; the computations use the unrounded cumulative indices, for example, 1.054725 for 2020. Because the BCC and slacks-based measures are invariant to rescaling any single variable across all units of a frontier, and within each annual frontier every monetary variable is observed in the same calendar year for all units, deflation leaves every efficiency score, ranking and classification unchanged. It affects only the monetary aggregates and projections reported in RMB. Re-estimating the network SBM on nominal data confirms this to within 2.0×10^{-10}.

4.2. Temporal Alignment Specifications and Matched Windows

Production-specification uncertainty concerns the technology representation; a second source of uncertainty concerns the temporal correspondence among research inputs, technological outputs and commercialization outcomes. Let t denote the starting year. A temporal specification $S\ell_Z\ell_C$ places research inputs in year t, technological output in $t + \ell_Z$ and commercialization outcomes in $t + \ell_Z + \ell_C$, with $\ell_Z, \ell_C \in \{0, 1, 2\}$. Table 4 lists the nine resulting specifications. The two principal alignments used throughout are the contemporaneous alignment $S00$ and the sequential alignment $S11$; we also report $S10$, which lags only the patent output, as an intermediate case.

Table 4. Temporal alignment specifications and analytical coverage.

Specification	Lag ℓ_Z	Lag ℓ_C	Total Span	Starting Years	Provinces	Observations
$S00$	0	0	0	2018–2023	28	168
$S10$	1	0	1	2018–2022	28	140
$S01$	0	1	1	2018–2022	28	140
$S11$	1	1	2	2018–2021	28	112
$S20$	2	0	2	2018–2021	28	112
$S02$	0	2	2	2018–2021	28	112
$S21$	2	1	3	2018–2020	28	84
$S12$	1	2	3	2018–2020	28	84
$S22$	2	2	4	2018–2019	28	56

Three matched windows are used. The main comparison between $S00$, $S10$ and $S11$ uses the 2018–2021 starting years, which are available under all three specifications and provide 112 province-year observations each. Specifications with a total span of three years are estimated on the smaller 2018–2020 window, which provides 84 observations each, and $S22$ retains only the 2018–2019 starting years and is reported for completeness. Statistics computed on different windows are not comparable and are never pooled. We retain the $S00$ observations for 2022 and 2023 for the full-period annual profiles and the numerical verification, but not for the matched specification comparisons, and we use no interpolation or extrapolation when constructing any temporal specification.

The design differs from dynamic DEA, which incorporates explicit intertemporal carry-over activities into a joint multi-period technology [31]. Here, observed annual variables are realigned and a separate provincial frontier is estimated for each starting year, so a frontier contains 28 DMUs; a window determines only which annual frontiers are summarized. In Section 3, $o \in \{1, \ldots, n\}$ indexes a province within a single annual frontier. In the multi-year summaries below, each observation is a province–starting-year pair: we write (o, t) where the starting year must be explicit, as in Equation (21), and let $o \in \mathcal{M}^{a,b}$ index matched province-year observations directly in Equation (20).

4.3. Sensitivity Metrics

For two specifications a and b evaluated on a matched set $\mathcal{M}^{a,b}$ of $m_{a,b}$ observations, the signed difference summarizes score sensitivity and the mean absolute difference (MAD):

$$\Delta_o^{a,b} = E_o^b - E_o^a, \qquad \mathrm{MAD}^{a,b} = \frac{1}{m_{a,b}} \sum_{o \in \mathcal{M}^{a,b}} |E_o^b - E_o^a|. \tag{20}$$

The signed difference captures the direction of the change, and the MAD captures its magnitude irrespective of direction. We evaluate rank stability with the maximum absolute rank change and with Spearman's rank correlation coefficient computed within each annual frontier [56]. We round scores to eight decimal places before assigning descending average ranks, so numerical solver noise does not determine the ordering of effectively tied observations. For the principal timing comparison, differences are defined as $S11 - S00$ and measure sensitivity to the alignment.

Classification comparisons report two rates on different denominators. The classification change rate uses observations determinate under both specifications, since a change can be observed only where both classifications are defined; the indeterminate share uses all matched observations, since it measures how often the interval in Proposition 1 fails to resolve.

4.4. Additional Design Variations

Further analyses vary design elements that the production and timing comparisons do not reach. Each is a complete re-estimation, and each is compared with the baseline on the same evaluated units:

- the stage weights, held at 0.5/0.5 in the baseline, are moved to 0.4/0.6 and to 0.6/0.4;
- contract value replaces realized income as the monetary output;
- the reference set of each annual frontier is extended from the 28 fixed provinces to every province whose records are strictly positive under both principal alignments for that starting year, which adds some of Xinjiang, Tibet and Qinghai in some years (30, 31 and 30 reference units for the starting years 2019, 2020 and 2021, and 28 for 2018; Tibet, for example, records no contracts in 2023 and is therefore not admissible under $S11$ for 2021); only the 28 fixed provinces are compared, giving $28 \times 4 \times 2 = 224$ pairs over the 2018–2021 window and the two principal alignments;
- in the years 2020–2022, in which all 31 provinces are admissible under $S00$, Tibet is removed from this extended reference set, and the remaining 30 provinces are compared, giving $30 \times 3 = 90$ pairs;
- the formation-lag specification $S10$ is run against $S00$;
- the intermediate link is redefined as a two-year and as a three-year rolling pool.

The baseline uses the annual flow of granted invention patents, whereas the technology available for commercialization is closer to a multi-year stock. The rolling-pool variation therefore redefines the link and re-estimates the whole model. It is a check on what the link measures: the pooled grants enter both as a Stage F output and as a Stage C input,

while research resources remain single-year flows. With ℓ_Z denoting the patent lag of the alignment under consideration, the two-year and three-year pools are

$$Z^{(2)}_{o,t} = \sum_{h=0}^{1} Z_{o,\,t+\ell_Z-h}, \qquad Z^{(3)}_{o,t} = \sum_{h=0}^{2} Z_{o,\,t+\ell_Z-h}. \tag{21}$$

The pool ends in the year the technological output is dated and extends backward, with research inputs and commercialization outcomes aligned exactly as in the baseline. The three link definitions are compared with $\ell_Z = \ell_C = 1$ on identical observations, on the 2019–2021 window (84 observations), which is the longest window admitting the three-year pool because that pool draws on patent data from the year before the starting year and on outcomes two years after it, and on the 2020–2021 window (56 observations). We hold every other element fixed, including the 28 provinces, the as-output link treatment and the equal stage weights. The stage weights enter $448 = 112 \times 2 \times 2$ pairs (two weight settings, two alignments), and the contract-value variation enters 224 pairs.

4.5. Implementation and Numerical Verification

All models are implemented in Python 3.12.11. We solve the linear programs with the HiGHS dual-simplex algorithm [57] under NumPy 2.4.1, SciPy 1.17.0 and pandas 3.0.0 [58–60], with primal and dual feasibility tolerances of 10^{-9} and otherwise SciPy's HiGHS defaults, with presolve enabled. For each network SBM evaluation, we solve three linear programs: the system program in Equation (10) and the two face programs in Equation (14). In the face programs, the defining equality of Equation (13) is imposed at the computed optimal value q^*_o after the objective row has been normalized by its largest coefficient, so the equality holds to within the primal feasibility tolerance. Residuals, gaps and link errors are measured in normalized units, with each variable divided by its largest value on the frontier.

Every logged linear program underlying the reported results passes the same checks. After each solve, we compute the primal residual and the duality gap, relative to the objective value, and stop the computation if either exceeds 2×10^{-7}. For network evaluations, we also verify the link-continuity equality, the two convexity equalities and the reconstruction identity in Equation (9). The checks cover all 20,002 logged linear programs: the principal solutions, all nine lag alignments, the regional frontiers, the stage-weight, contract-value, reference-set, Tibet, pooled-link and nominal-data variations, the conditional programs with their income bounds, the relaxed face programs, the re-estimations of Section 5.5 and the solver cross-checks; Appendix B gives the counts by analysis. All of them pass, with a largest primal residual of 9.8×10^{-10} and a largest relative duality gap of 9.9×10^{-11}. Across the 1960 programs of the principal solutions, the largest primal residual is 5.6×10^{-10}, the largest duality gap is 9.9×10^{-11} and the largest link-continuity error is 1.2×10^{-10}. Re-solving the principal programs with the HiGHS interior-point method gives the same system efficiencies to within 2.9×10^{-13}.

Because the face equality holds only to the primal feasibility tolerance, the computed face bounds are exact only up to that tolerance. Computed endpoint differences of order 10^{-10} or smaller lie below the feasibility tolerance and are read as numerically degenerate rather than as proof of exact uniqueness; a decomposition is treated as numerically non-unique when the width of either stage-efficiency interval in Equation (15) exceeds 10^{-6}, and the classification tolerance in Equation (16) is $\varepsilon = 10^{-7}$. The tolerance guards only against numerical noise and is not an economic indifference threshold. Because $B \geq 1$, an error η in a stage expansion changes the stage efficiency $1/B$ by at most η; $\varepsilon = 10^{-7}$ therefore lies two orders of magnitude above the solver feasibility tolerance of 10^{-9} and one order below the threshold of 10^{-6} used to flag numerically non-unique intervals. Section 5.4 shows that the reported classifications are unchanged over a wide range of tolerances

around this choice. To check that imposing the equality at the computed q_o^* does not itself collapse the face, the face programs were also solved with the equality relaxed to $w_F B_{F,o} + w_C B_{C,o} \geq (1-\delta) q_o^*$ for $\delta = 10^{-8}$ and 10^{-6}.

We finally verify the uniqueness of the stage decomposition in exact rational arithmetic, independently of solver tolerances. For each principal system program, we read a basis from the dual-simplex solution, supplement it at degenerate vertices with the columns with the smallest reduced costs, and recompute the basic solution, the dual vector and all reduced costs using exact rational numbers from the stored data. A primal feasible basis with all reduced costs nonpositive proves optimality and gives q_o^* exactly. If every nonbasic reduced cost is strictly negative, the optimal solution, and hence the stage decomposition, is unique [54]. If $q_o^* = 1$ exactly, the decomposition is unique by Remark 5; otherwise, $B_{F,o}$ is maximized and minimized over the optimal face defined by the exact reduced costs, and the optimal bases of both programs are again verified exactly. Supplementary Data S1 lists, for each principal program, the eight basic variables of the certified basis of Equation (10) written in equality form, each variable named as the Stage F or Stage C intensity of a province or as the slack of one of the five inequality rows, together with the exact optimal value q_o^* as a rational number. Each certificate can therefore be checked from the data without an LP solver.

As a second implementation, we also solved the network SBM specification with DEAPack 2.0.1, a separately written model compiler, under the same output orientation, VRS assumption, stage weights, and as-output link treatment [61]. The two implementations return the same efficiency scores for all 280 province-year network SBM solutions of the two principal alignments (168 under *S*00, covering 2018–2023, and 112 under *S*11, covering 2018–2021), with a maximum absolute difference of 2.9×10^{-13}. The DEAPack stage decomposition lies inside the optimal-face interval in every case. Both implementations call the SciPy interface to HiGHS, so this comparison verifies the model formulation rather than the solver.

5. Results

5.1. Sample Characteristics and Annual Profiles

Table 5 reports descriptive statistics for the six variables over the 168 province-year observations of the fixed sample.

Table 5. Descriptive statistics of the analytical sample.

Variable	Obs.	Mean	Std. Dev.	Min.	Median	Max.
H	168	13,697.52	9845.26	894	11,437.50	60,492
R	168	4,185,005.48	4,497,935.21	151,217	2,435,296.70	26,747,104.35
Z	168	4699.18	4726.01	74	3178.50	25,592
N	168	806.45	1091.30	3	469.50	8018
I	168	153,944.67	174,572.44	245	79,914.65	811,568.46
A	168	389,389.64	575,836.75	250	176,815.94	3,793,969.66

The variables span several orders of magnitude, and the output mix differs widely across units: contracts range from 3 to 8018 and realized income from 245 to 811,568 thousand RMB. Over 2018–2021, the ratio of realized income to R&D expenditure has a province median of 2.9% and ranges from 0.3% in Yunnan to 11.7% in Chongqing. This heterogeneity in the proportions of the two commercialization outputs is the setting in which radial and slacks-based adjustments can diverge.

Figure 2 reports mean stage and system efficiency by starting year. Each point is the mean over the 28 provinces of that year's own frontier, so the series compares annual frontiers rather than a single pooled frontier, and the shaded bands give the interquartile range across

provinces. Under *S*00, mean Stage F efficiency increases from 0.667 in 2018 to 0.762 in 2022 and is 0.715 in 2023, whereas mean Stage C efficiency never exceeds 0.420. The Stage C-to-Stage F ratio therefore lies between 0.44 and 0.55 each year, and mean system efficiency lies between 0.376 in 2018 and 0.494 in 2022. Under *S*11, the stage order is unchanged, with Stage C between 0.333 and 0.419 and system efficiency between 0.405 and 0.506.

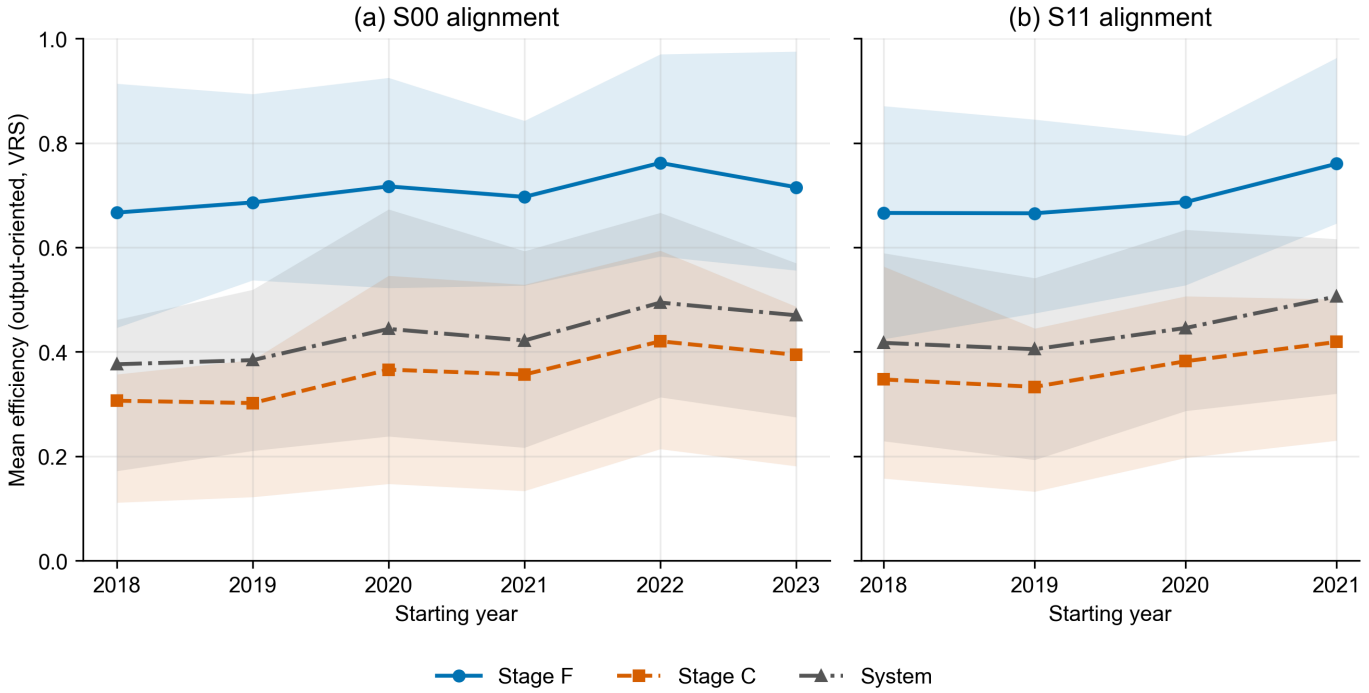


Figure 2. Mean stage and system efficiency by starting year.

The stage gap is not specific to a single year but appears in all ten annual frontiers of *S*00 and *S*11. Individual provincial scores are volatile, however, because each frontier is estimated from 28 observations, as Figure 3 shows for the complete province-by-year matrix; each cell is a province-year score against that year's own frontier, rows are ordered by the mean efficiency across both alignments over 2018–2021, and the two panels share that order. For example, Ningxia is recorded at 0.08 in 2018 and 1.00 in 2021 under *S*00, and Beijing at 0.51 in 2019 and 1.00 in 2020. Annual scores should therefore be read against their own frontier, while multi-year averages summarize the recurring pattern.

5.2. Sensitivity to the Production Representation

Table 6 reports mean efficiency scores under the three production specifications in Panel A, the corresponding score and ranking sensitivity within each alignment in Panel B, and the sensitivity of the alignment itself in Panel C. All three panels refer to the common 2018–2021 window, in which each specification provides 112 province-year observations; changes are averaged over the four annual frontiers, the mean change is the signed average difference and the MAD its absolute counterpart, and the minimum rank correlation is the lowest of the four annual values. Panel A reports *S*10 over the same four years as *S*00 and *S*11, although *S*10 is estimated for 2018–2022.

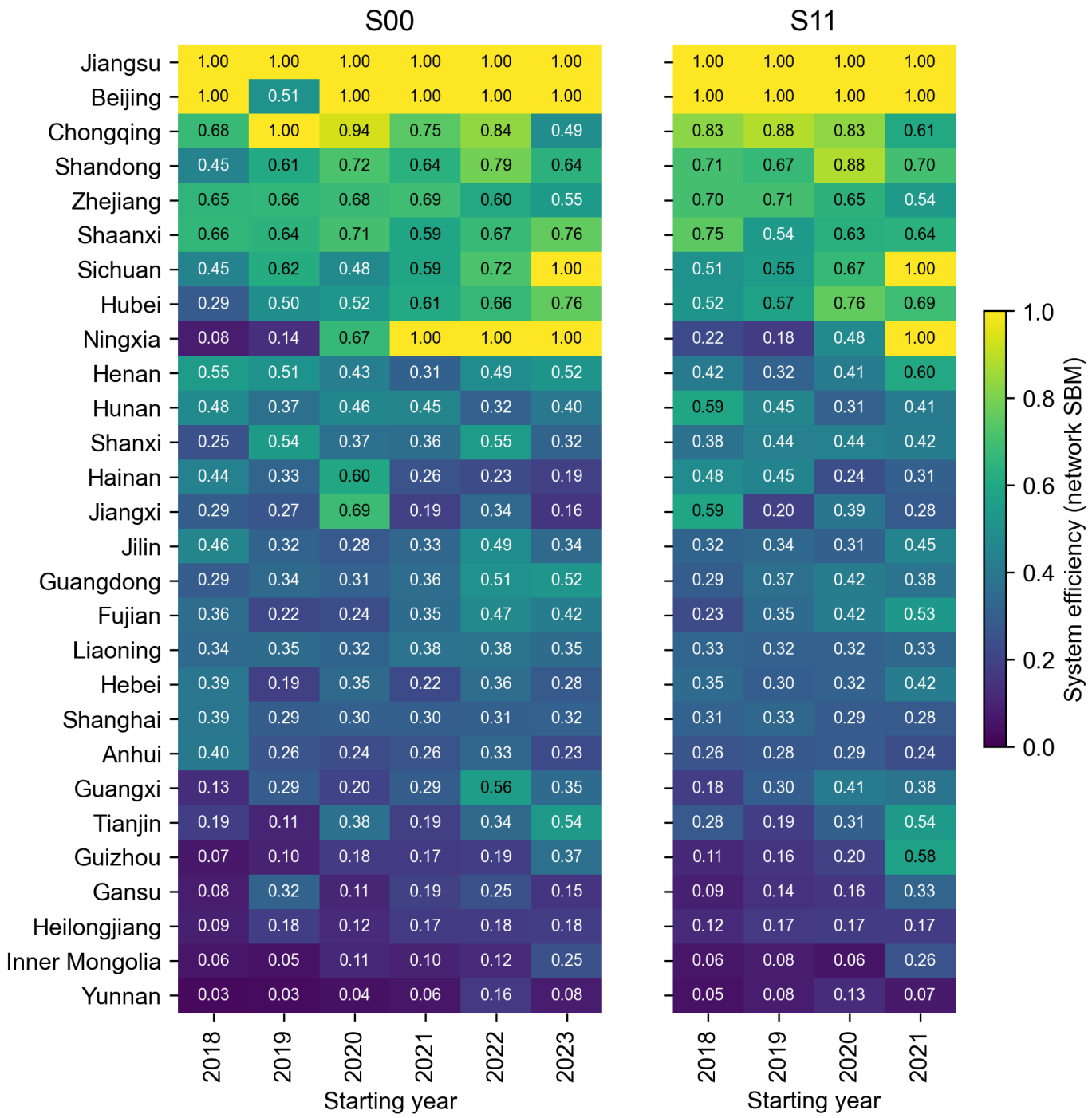


Figure 3. System efficiency of the network SBM by province and starting year: contemporaneous (*S*00, left) and sequential (*S*11, right) alignment.

The panels separate model-determined features from empirical features. First, the BCC and independent SBM scores for Stage F coincide exactly because of the single-output structure stated in Remark 1. Moving from the independent SBM to the network SBM raises mean Stage F efficiency by 0.0047 under *S*00, 0.0051 under *S*10 and 0.0115 under *S*11. In each alignment, the MAD equals the mean change, as Remark 3(b) requires, since no observation can lose Stage F efficiency in this step; individual ranks nevertheless move by up to 9, 14.5 and 21 positions.

Table 6. Mean efficiency scores and sensitivity to production-technology specifications. In Panel B, Mean is the mean under the first specification of each contrast, and changes are the second minus the first; in Panel C, Mean is the mean under the row alignment, and changes are the row alignment minus the comparison alignment. Ranks are averaged within each annual frontier so that rank changes can be half-integers.

Panel A. Mean efficiency scores, common 2018–2021 window

	Stage F (Technology Formation)			Stage C (Commercialization)			
Timing	**BCC**	**SBM**	**Network**	**BCC**	**SBM**	**Network**	**System**
*S*00	0.6868	0.6868	0.6915	0.5507	0.4408	0.3327	0.4065
*S*10	0.6830	0.6830	0.6881	0.5615	0.4492	0.3449	0.4185
*S*11	0.6830	0.6830	0.6945	0.5879	0.4813	0.3703	0.4435

Panel B. Score and ranking sensitivity within each alignment

				Change		Rank Stability	
Timing	**Stage**	**Contrast**	**Mean**	**Mean Chg.**	**MAD**	**Min. Annual ρ_s**	**Max. Rank Chg.**
*S*00	F	SBM/Net	0.6868	0.0047	0.0047	0.9755	9
*S*00	C	BCC/SBM	0.5507	−0.1099	0.1099	0.7492	19
*S*00	C	SBM/Net	0.4408	−0.1081	0.1081	0.6439	23
*S*00	C	BCC/Net	0.5507	−0.2180	0.2180	0.4829	23
*S*10	F	SBM/Net	0.6830	0.0051	0.0051	0.9384	14.5
*S*10	C	BCC/SBM	0.5615	−0.1122	0.1122	0.7492	19
*S*10	C	SBM/Net	0.4492	−0.1044	0.1044	0.8090	19.5
*S*10	C	BCC/Net	0.5615	−0.2166	0.2166	0.4938	20
*S*11	F	SBM/Net	0.6830	0.0115	0.0115	0.8075	21
*S*11	C	BCC/SBM	0.5879	−0.1065	0.1065	0.7606	16
*S*11	C	SBM/Net	0.4813	−0.1111	0.1200	0.7205	22
*S*11	C	BCC/Net	0.5879	−0.2176	0.2196	0.6184	22

Panel C. Sensitivity to temporal alignment (network SBM, 112 observations)

				Change		Rank Stability	
Alignment	**Stage**	**Contrast**	**Mean**	**Mean Chg.**	**MAD**	**Min. Annual ρ_s**	**Max. Rank Chg.**
*S*10	System	versus *S*00	0.4185	0.0120	0.0970	0.7702	16
*S*10	System	versus *S*11	0.4185	−0.0250	0.0759	0.8662	14
*S*11	System	versus *S*00	0.4435	0.0370	0.0974	0.7820	15

Second, Stage C is considerably more specification-dependent. Replacing radial output expansion with slacks-based adjustment lowers mean Stage C efficiency by 0.1099 under *S*00, 0.1122 under *S*10 and 0.1065 under *S*11; the sign is fixed by Remark 3(a), so the MAD again equals the absolute mean change, and only the magnitude is empirical. Imposing the linked network structure further lowers mean Stage C efficiency by 0.1081, 0.1044 and 0.1111. The sign of this step is not determined by the models (Remark 3(c)): under *S*00 and *S*10, no observations gain, but under *S*11, the MAD of 0.1200 exceeds the absolute mean change of 0.1111, so some observations gain. The decline accompanies the move to joint system optimization with a common benchmark link flow, and the comparison does not separate the contribution of the benchmark link flow from that of the optimization structure. The benchmark link flow is nonetheless substantial (Remark 4): under *S*00, the benchmark link flow exceeds the observed patent count in 91 of the 112 observations, and the median ratio of the benchmark to the observed patent count over the 112 observations is 1.51 (1.67 among the 91 with expansion). The combined BCC-to-network-SBM difference is 0.2180, 0.2166 and 0.2176, and the corresponding rank correlations fall to 0.4829, 0.4938 and 0.6184

in the least favorable year. Linking the two stages also contracts the efficient set: of the 112 matched observations, 20 lie on the Stage C frontier under BCC but only nine under the network SBM, with 22 and 10 under *S*11.

Third, alignment matters less than the production specification, but not negligibly. Panel C compares the three alignments on the same 112 observations. Relative to contemporaneous alignment, *S*10 raises mean system efficiency by 0.0120 and *S*11 by 0.0370; the two differ by −0.0250. Score sensitivity is of the same order in both comparisons with *S*00 (MADs of 0.0970 and 0.0974), and the MAD between *S*10 and *S*11 is 0.0759. *S*10 and *S*11 agree more closely (lowest annual rank correlation: 0.8662) than either does with *S*00 (0.7702 and 0.7820), so the main change in both scores and rankings occurs between contemporaneous and lagged treatment rather than between the two lag structures.

The specification effect differs across provinces. Figure 4 shows mean Stage C efficiency under the three specifications for each province under contemporaneous alignment, averaged over 2018–2021; provinces are ordered by their BCC mean. When radial expansion is replaced by slacks-based adjustment, mean Stage C efficiency falls by 0.314 in Gansu, 0.246 in Hebei, 0.195 in Anhui, 0.195 in Inner Mongolia and 0.194 in Shanghai, whereas Shaanxi (0.020), Sichuan (0.008), Chongqing, Ningxia and Jiangsu are essentially unaffected; provinces whose contract numbers and realized income fall short by different proportions are the most exposed to this choice. The largest BCC-to-network difference, 0.557 for Ningxia, illustrates Remark 4: Ningxia's Stage C efficiency is one under both independent specifications in each of the four starting years, whereas its network Stage C efficiencies are 0.046, 0.101, 0.626 and 1.000, with benchmark link flows of 2.9, 4.4, 1.4 and 1.0 times its observed patent count.

5.3. Sensitivity to Temporal Alignment

Table 7 compares *S*00, *S*10 and *S*11 by starting year over 2018–2021; Table 8 places them in the complete grid in which the patent lag ℓ_Z and the commercialization lag ℓ_C each take the values zero, one and two years, and Figure 5 reports the MAD of each grid specification from contemporaneous alignment.

Table 7. Sensitivity of system efficiency to temporal alignment by starting year, 2018–2021 (28 observations per year). The rank-correlation and rank-change columns report the more extreme of the *S*10 vs. *S*00 and *S*11 vs. *S*00 comparisons; the classification change refers to *S*11 vs. *S*00.

Year	Mean $S00$	Mean $S10$	Mean $S11$	MAD $S10 - S00$	MAD $S11 - S00$	Min. Annual ρ_s	Max. Rank Change	Class. Change
2018	0.3763	0.3737	0.4172	0.1125	0.0888	0.7702	16	3.6%
2019	0.3842	0.4055	0.4051	0.0970	0.0847	0.8214	12	7.1%
2020	0.4439	0.4080	0.4457	0.0886	0.1023	0.8067	14	3.6%
2021	0.4218	0.4868	0.5062	0.0900	0.1139	0.7820	15	10.7%

The *S*11-to-*S*00 comparison gives a mean difference of 0.0370 and a MAD of 0.0974. Province-specific changes occur in both directions: 69 of the 112 matched observations are more efficient under *S*11, 35 are less efficient, and 8 are unchanged within tolerance. For that pairing, annual rank correlations range from 0.782 to 0.877, and the maximum rank change is 15 positions. *S*10 lies between the other two alignments in mean score but not in distance: relative to *S*00, it produces MADs of 0.0886–0.1125, rank correlations of 0.770–0.875 and maximum rank changes of 10–16 positions, whereas relative to *S*11, the corresponding ranges are 0.0605–0.0915, 0.866–0.909 and 9–14 positions.

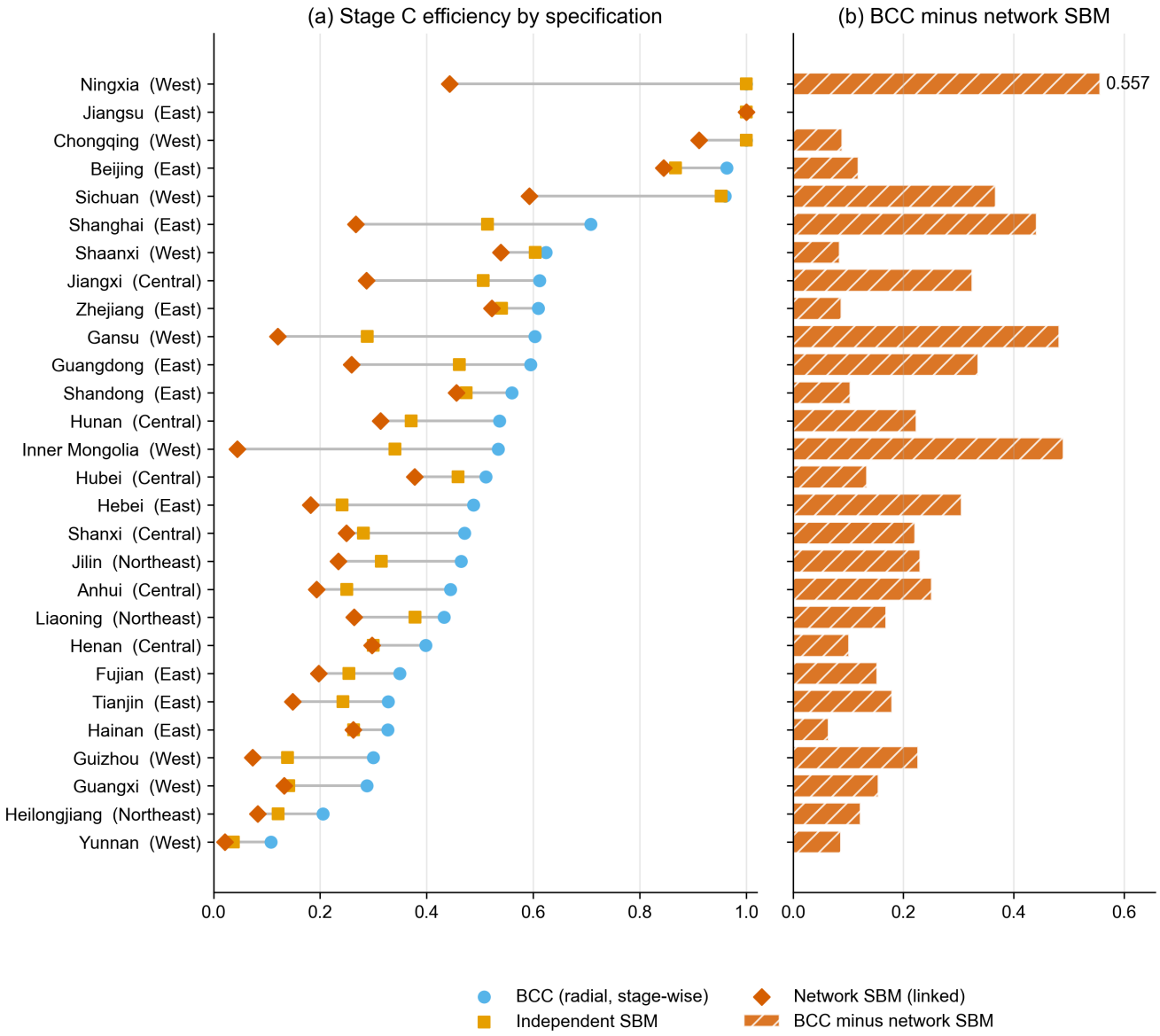


Figure 4. Stage C efficiency under the three production-technology specifications, contemporaneous alignment ($S00$), 2018–2021 province means: (**a**) scores under each specification; (**b**) BCC minus network SBM.

The lag grid shows that sensitivity varies with the total span but not monotonically. In the 2018–2020 window, the MAD from $S00$ is 0.0867–0.0994 for one-year spans, 0.0879–0.0971 for two-year spans, and 0.1044–0.1146 for three-year spans; $S22$ gives 0.1201 on its own 2018–2019 window. Because $S22$ does not have the 2018–2020 window, the comparison across spans mixes windows, so Table 8 repeats it on the 2018–2019 window that all nine specifications support. There, the MADs are 0.0953–0.1048 at span one, 0.0868–0.0874 at span two, 0.0961–0.0969 at span three, and 0.1201 at span four. The non-monotonic pattern is preserved and is sharper on the common window, where the two-year span produces the smallest differences, so the ordering does not describe a monotone span effect.

Table 8. Sensitivity of system efficiency across the full lag grid. Each annual frontier is estimated separately; columns 5–10 summarize the starting years 2018–2020 for *S*00–*S*21 and 2018–2019 for *S*22, whose two-year commercialization lag excludes 2020, and the MAD of *S*22 is computed against *S*00 on the same 56 observations. The last two columns repeat the comparison on the 2018–2019 window, which all nine specifications support (56 observations each).

				Window 2018–2020 (*S*22: 2018–2019)						Common 2018–2019	
Spec.	ℓ_Z	ℓ_C	**Span**	**Obs.**	**Mean E^{NET}**	**MAD vs. *S*00**	**Min. ρ_s**	**Max. Rank Chg.**	**Stage C Weaker (%)**	**MAD**	**Stage C Weaker (%)**
*S*00	0	0	0	84	0.4014	0.0000	1.0000	0	86.9	0.0000	87.5
*S*01	0	1	1	84	0.4067	0.0867	0.8088	13	85.7	0.0953	85.7
*S*02	0	2	2	84	0.4362	0.0879	0.8248	12	84.5	0.0872	85.7
*S*10	1	0	1	84	0.3957	0.0994	0.7702	16	86.9	0.1048	87.5
*S*11	1	1	2	84	0.4227	0.0919	0.8122	14	83.3	0.0868	83.9
*S*12	1	2	3	84	0.4401	0.1044	0.7827	15	84.5	0.0969	83.9
*S*20	2	0	2	84	0.4264	0.0971	0.7986	15	84.5	0.0874	85.7
*S*21	2	1	3	84	0.4443	0.1146	0.7279	14	84.5	0.0961	82.1
*S*22	2	2	4	56	0.4338	0.1201	0.6901	16	82.1	0.1201	82.1

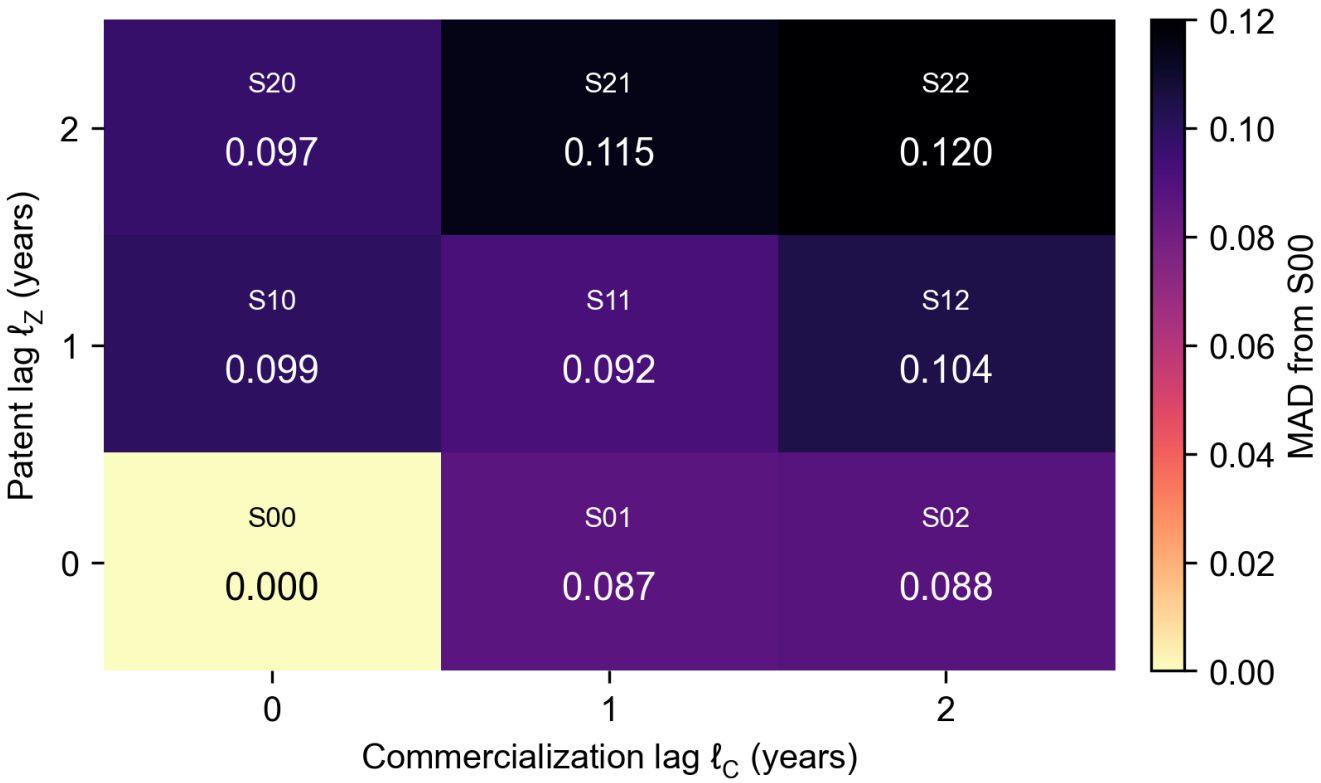


Figure 5. MAD from contemporaneous alignment across the lag grid. Cells with a total span of at most three summarize the starting years 2018–2020; the *S*22 cell summarizes 2018–2019 and is compared with *S*00 on the same 56 observations.

Among specifications with an identical two-year span, allocating the whole lag to the patent side (*S*20: MAD 0.0971 on the 2018–2020 window, lowest rank correlation 0.7986) changes results more than allocating it to the commercialization side (*S*02: 0.0879 and 0.8248), with the balanced specification closest to *S*02 (*S*11: 0.0919 and 0.8122). On the common window, *S*11 gives 0.0868, below *S*02 at 0.0872 and *S*20 at 0.0874: the order of the two polar placements is preserved, but the gap between them narrows from 0.0092 to

0.0002, so the window changes the margin and the position of *S*11 relative to them, not the order of the polar placements. These differences are small and are not tested for statistical significance, so they do not support a mechanism-specific interpretation.

Across the grid, commercialization remains the weaker stage in 82.1–86.9% of observations in the main columns of Table 8 (83.3–86.9% for the eight specifications on the 2018–2020 window) and in 82.1–87.5% on the common window; the minimum rank correlation is 0.690, and the largest rank movement is 16 positions. Individual classifications move, but no observation has a numerically non-unique stage decomposition under the stated tolerances. The principal pairing *S*00–*S*11 lies within the range of the grid.

5.4. Stage Classification and Optimal-Face Bounds

By Proposition 1, the two auxiliary programs bound the stage gap exactly, so the network SBM classifications in Table 9 hold over the entire system-optimal face up to the tolerances of Section 4.5. For the BCC and independent SBM specifications, the stage scores are unique optimal values, and we apply the classification to the point gap.

Table 9. Weaker stage classifications under alternative production-technology specifications, 112 matched observations per column. Network SBM classifications are evaluated over the system-optimal face (Proposition 1).

Classification	*S*00			*S*11		
	BCC	**Indep. SBM**	**Network**	**BCC**	**Indep. SBM**	**Network**
Stage C weaker	71	82	96	73	80	93
Stage F weaker	34	23	7	33	26	9
Equal	7	7	9	6	6	10
Indeterminate	0	0	0	0	0	0

The number of observations in which the weaker stage is Stage C rises from 71 under BCC to 82 under the independent SBM and to 96 under the network SBM in the *S*00 window (63.4%, 73.2% and 85.7% of 112), with 73, 80 and 93 (65.2%, 71.4% and 83.0%) under *S*11. The predominant classification is the same under all six combinations, but its frequency depends on the benchmark link flow at which commercialization is evaluated. Conditional on the observed patent count (independent SBM), Stage C is the weaker stage in 73.2% and 71.4% of observations; at the benchmark link flow (network SBM), in 85.7% and 83.0% of observations. The difference reflects joint changes in the optimization structure and the treatment of the benchmark link flow; it cannot be attributed solely to the higher benchmark link flow. Remark 4 accounts for the composition of the network classification. Under *S*00, the link expands in 91 observations: 84 are classified as Stage C weaker and seven as Stage F weaker; of the 21 observations without expansion, 12 are Stage C weaker, and nine are on the frontier, with both stages equal to 1. Under *S*11, the corresponding counts are 92 (83 and nine) and 20 (10 and 10). Every Stage F weaker classification occurs with an expanded link, and no unexpanded observation is classified as Stage F weaker. Therefore, part of the network frequency is structural. A second structural feature is the construction of the two indices: the gap compares a single output ratio with an average of two output ratios, and under *S*00, the median Stage C expansion is 3.9, the largest is 74.8, and 18 of the 112 observations have expansions above ten. Section 5.5 examines how this construction affects the classification.

The optimal-face analysis shows that the specification-driven spread does not arise from non-uniqueness of the stage decomposition. Across the 280 principal solutions, the computed endpoints of the Stage F efficiency interval compatible with the system optimum differ by at most 2.4×10^{-10} in absolute value, and those of Stage C by at most 1.3×10^{-11} (some computed widths are slightly negative because of floating-point error); both lie be-

low the feasibility tolerance, no observation is flagged as numerically non-unique, and the interval-based classification coincides with the classification obtained from the solver-selected decomposition in every case. Three numerical checks support this result. First, when the face equality is relaxed by $\delta = 10^{-8}$ and 10^{-6}, the largest stage-efficiency interval widens to 2.4×10^{-8} and 2.4×10^{-6}, in proportion to δ, and none of the 280 classifications changes, which is the behavior expected when the optimal face is a single point in $(B_{F,o}, B_{C,o})$ up to numerical tolerance. Second, the classifications lie far from the tolerance: among observations classified as Stage C or Stage F weaker, the smallest $|G_o|$ is 0.017 under $S00$ and 0.0038 under $S11$, against $\varepsilon = 10^{-7}$, while in the equality class, $|G_o|$ is at most 5.82×10^{-13}. Consequently, every ε between 6.0×10^{-13} and 3.8×10^{-3} yields exactly the reported network classifications. The same holds for every classification reported in this paper, including the BCC and independent SBM classifications, the lag grid, the regional frontiers, the robustness variations and the checks of Section 5.5, for every ε in $[1.1 \times 10^{-9}, 1.6 \times 10^{-3}]$ (Table A4 in Appendix C). With $\varepsilon = 0$, by contrast, numerical noise of order 10^{-13} would relabel eight of the nine $(S00)$ and all 10 $(S11)$ observations whose stage efficiencies are equal in exact arithmetic, so a positive tolerance is needed. Larger values act as economic indifference bands: with $\varepsilon = 0.05$, the share of network observations labeled Stage C weaker is 85.7% under $S00$ and 82.1% under $S11$, and the remaining changes move observations from a weaker-stage label to equal. Third, the procedure does detect non-uniqueness when it is present, as Example 1 shows. The exact certificates settle the question for the data as stored: For 267 of the 280 principal system programs, every nonbasic reduced cost is strictly negative, so the optimal solution is unique. Of the remaining 13, which are dual degenerate, 11 are efficient with $q_o^* = 1$ exactly, and for the last two, $B_{F,o}$ takes the same value at its maximum and minimum over the optimal face. The stage decomposition is therefore unique for all 280 principal solutions.

The degeneracy is consistent with Remark 5, under which a non-trivial interval requires the benchmark link to vary along an exactly flat direction of the objective. For this sample, the procedure confirms the solver-selected decompositions, and the remaining risk to the stage diagnosis lies in the production specification.

The uniqueness of the stage pair does not imply uniqueness of the output mix, so we examine the income projection in Equation (17) separately. The exact certificates settle this as well: the income projection is unique for all 280 principal solutions, because the optimal solution is unique in 267 cases, projected income equals observed income when $q_o^* = 1$ (the 11 dual-degenerate efficient cases), and the exact maximum and minimum of $I\boldsymbol{\lambda}_C$ over the optimal face coincide in the remaining two. Over the 112 $S00$ observations of the main window, the aggregate income shortfall is therefore RMB 17.00 billion at every optimal solution.

Temporal alignment changes seven of the 112 network SBM classifications, including four direct reversals: Jiangxi in 2018 and Ningxia in 2020 shift from a weaker Stage C to a weaker Stage F, Hubei in 2021 changes in the same direction, and Chongqing in 2021 changes in the opposite direction. Aggregate counts remain stable: commercialization is the weaker stage in 85.7% of observations under $S00$ and 83.0% under $S11$, with stage equality in 8.0% and 8.9% and a weaker Stage F in 6.3% and 8.0%.

5.5. Construction of the Stage Measures and Small Denominators

The two stage measures are constructed differently. Because $G_o = 1/B_{F,o} - 1/B_{C,o}$, the gap G_o is positive exactly when $B_{C,o} > B_{F,o}$, that is, when the average of two commercialization ratios, $B_{C,o} = \frac{1}{2}(r_{N,o} + r_{I,o})$ with $r_{N,o} = N\boldsymbol{\lambda}_C/N_o$ and $r_{I,o} = I\boldsymbol{\lambda}_C/I_o$, exceeds the single formation ratio $r_{Z,o} = Z\boldsymbol{\lambda}_F/Z_o$. Small denominators can inflate any of the three ratios, so they do not by themselves favor either label; a small patent count also inflates $B_{F,o}$. The asymmetry lies in the averaging. Since $B_{C,o} \geq \frac{1}{2}\max\{r_{N,o}, r_{I,o}\}$, one extreme commercialization

ratio can raise $B_{C,o}$ above $B_{F,o}$ even when the other ratio is below it, and the label then depends on the equal weighting of contracts and income rather than on a shortfall common to both outputs. Under *S*00, the median ratios are 1.51 for patents, 3.47 for contracts and 3.52 for income.

Table 10 reports four checks on the principal alignments. First, because the stage decomposition and the income projection are unique (Section 5.4), the ratios $r_{N,o}$ and $r_{I,o}$ are unique as well. For 78 of the 96 "Stage C weaker" labels under *S*00 and 71 of the 93 under *S*11, both commercialization ratios individually exceed the formation ratio, so these labels hold for every convex weighting of contracts and income at the reported benchmark; the remaining 18 and 22 labels depend on the equal weighting. Second, observations with small denominators are labeled "Stage C weaker" more often: 93.8% (*S*00) and 87.5% (*S*11) of the 16 observations with one of the three smallest contract or income values of their annual frontier, against 84.4% and 82.3% of the others. However, 93.3% and 92.9% of these labels satisfy the ratio-dominance condition, so they reflect shortfalls in both outputs rather than one inflated ratio. Third, removing small-denominator observations from both the evaluated and the reference sets changes few classifications. Without the 18 (*S*00) and 13 (*S*11) observations whose Stage C expansion exceeds ten, one and none of the remaining classifications change; without the 16 observations that have one of the three smallest contract or income values of their annual frontier, four change under each alignment, all from "Stage C weaker" to "equal", because these units reach the smaller frontier. Fourth, re-estimating the network SBM with a single Stage C output removes the difference in construction. Commercialization remains the predominant weaker stage, with shares of 90.2% and 88.4% when contracts are the only output and 83.9% and 79.5% when income is. The predominant classification is therefore not produced by a few small-denominator observations or by the averaging of two ratios, although its frequency varies by up to 5.4 percentage points relative to the baseline on the same observations across these checks.

Table 10. Stage-measure construction and small denominators: observations labeled "Stage C weaker" by the network SBM, 2018–2021.

	***S*00**		***S*11**	
Analysis	**Obs.**	**Stage C Weaker**	**Obs.**	**Stage C Weaker**
Baseline	112	85.7%	112	83.0%
Stage C weaker labels with $\min(r_N, r_I) > r_Z$	96	78 (81.3%)	93	71 (76.3%)
Three smallest *N* or *I* in the annual frontier	16	93.8%	16	87.5%
Other observations	96	84.4%	96	82.3%
Seven smallest *N* or *I* in the annual frontier	33	97.0%	32	93.8%
Other observations	79	81.0%	80	78.8%
Re-estimated without $B_{C,o} > 10$	94	81.9% (1)	99	80.8% (0)
Re-estimated without the three smallest *N* or *I*	96	80.2% (4)	96	78.1% (4)
Stage C output: contracts only	112	90.2% (5)	112	88.4% (12)
Stage C output: income only	112	83.9% (8)	112	79.5% (12)

Notes: r_Z, r_N and r_I are the ratios of the benchmark to the observed patents, contracts and income at the stage decomposition and income projection. Re-estimations remove the observations from both the evaluated and the reference sets of each annual frontier. Numbers in parentheses are classification changes relative to the baseline on the same observations; in the two leave-out rows all changes are from "Stage C weaker" to "equal".

5.6. Reference-Set Dependence and Frontier Discrimination

Under the national frontier, regions differ less in mean scores than in the composition of their classifications. Figure 6 reports, for the 2018–2021 window, mean Stage F and Stage C efficiency by region in panel (a) and each region's classification composition as a percentage of its own observations in panel (b); the group sizes are 40 province-year observations for the eastern region, 36 for the western, 24 for the central and 12 for the northeastern. Mean system efficiency is highest in the eastern region (0.500 under *S*00, 0.549 under *S*11) and lowest in the northeastern region (0.279 under both), with the central (0.400, 0.427) and western (0.350, 0.393) regions between them. Stage C lies below Stage F in every region. The eastern region records stage equality in 17.5–20.0% of observations, the central and northeastern regions record a weaker Stage C in every observation under *S*00, and the western region is the only one in which technology formation is the weaker stage for a non-trivial share of observations. The eastern region leads the western region in both stages, with Stage C averaging 0.414 against 0.320 and Stage F 0.755 against 0.624.

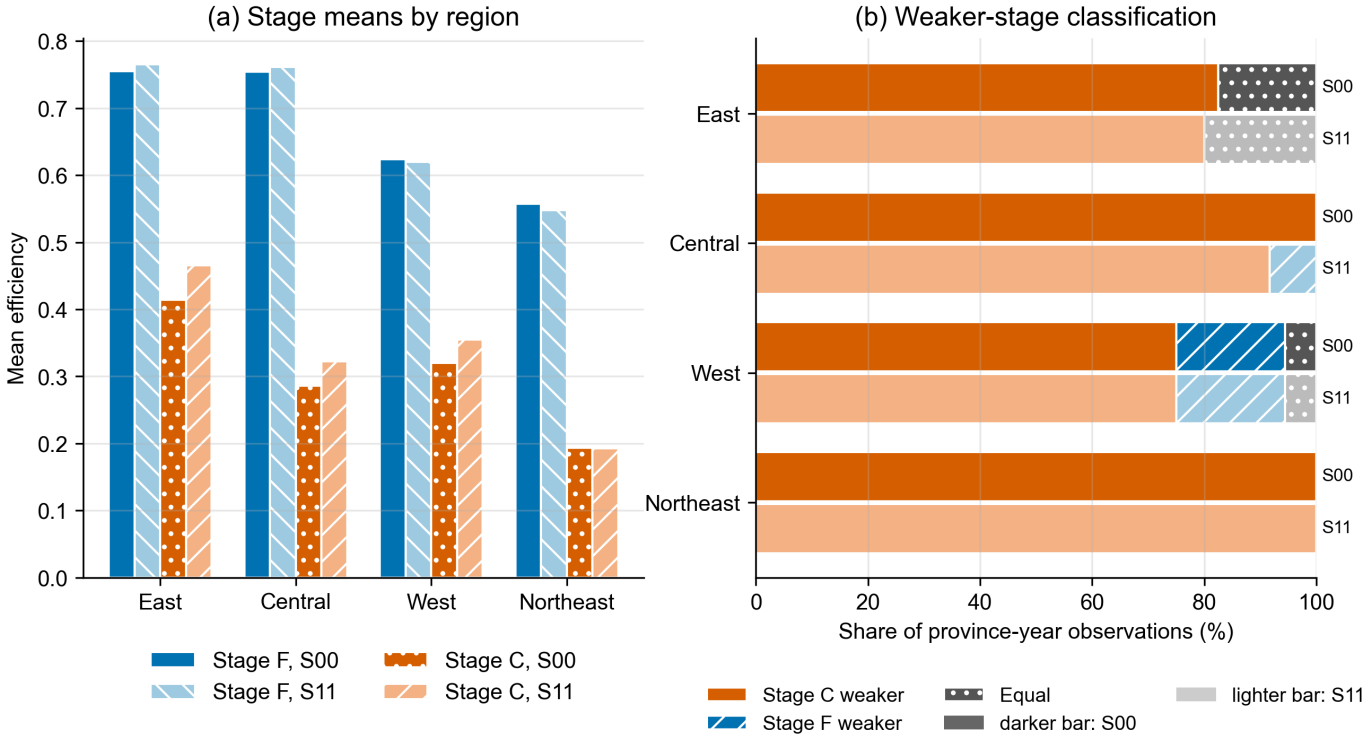


Figure 6. Regional stage means and weaker stage classifications against the national frontier, 2018–2021: (**a**) mean Stage F and Stage C efficiency; (**b**) classification shares. In both panels, darker bars refer to *S*00 and lighter bars to *S*11.

To test whether these differences reflect technology or the reference set, we replace the national frontier with a regional one, re-benchmarking each province against the provinces in its own region on the *S*00 window. By Remark 6, the system efficiency cannot fall when the reference set shrinks. Table 11 reports the ratio $\Gamma_{o,k}$ of Equation (19) and the resulting classifications.

All 112 observations have stage ratios at or below one, with the largest at 1.000; Remark 6 guarantees this only for the system efficiency, so for the stage scores, it is a property of this sample.

Table 11. Stage diagnosis under the national frontier and under four regional reference sets (*S*00, 2018–2021). Equal coincides with $E^{NET} = 1$ against the corresponding frontier, which is output-oriented efficiency rather than Pareto–Koopmans efficiency; percentages are rounded and may not sum to 100.

Reference Set	Units	Obs.	Median $\Gamma_{o,F}$	Median $\Gamma_{o,C}$	Stage C Weaker (%)	Stage F Weaker (%)	Equal (%)	Indet. (%)	Agree (%)
National	28	112	–	–	85.7	6.3	8.0	0	–
Regional: East	10	40	1.000	0.858	72.5	0.0	27.5	0	90.0
Regional: West	9	36	0.781	0.681	55.6	2.8	41.7	0	63.9
Regional: Central	6	24	0.823	0.376	37.5	4.2	58.3	0	37.5
Regional: Northeast	3	12	0.574	0.250	33.3	0.0	66.7	0	33.3

The share of observations on the regional frontier rises as the region shrinks: 27.5% in the east (ten provinces), 41.7% in the west (nine), 58.3% in the central region (six) and 66.7% in the northeast (three). In this sample, the equality class and the on-frontier class coincide exactly. All 48 observations classified as stage equality under the regional frontiers score exactly 1.000 on both stages, and no observation is on a frontier without being classified as equal, as Remark 2(b) implies for the latter direction. The equality column of Table 11 therefore reads directly as the share of the region that the regional frontier cannot separate.

The classifications change accordingly. Of the 112 observations, the national frontier records a weaker Stage C in 96 and a weaker Stage F in 7. Under the regional frontiers, the weaker-Stage-C count falls to 62, and the weaker-Stage-F count to two, and the two classifications agree in 72 observations (64.3%). Almost all movement is into stage equality rather than between the two stages: 33 observations the national frontier calls a weaker Stage C, and six of the seven it calls a weaker Stage F, become stage equality, while only one observation crosses directly from one weaker stage to the other. Together with the nine observations already classified as equal under the national frontier, these 39 observations make up the 48 equality observations under the regional frontiers. The reclassification therefore concentrates where the regional benchmark has least discriminating power, and it cannot be read as evidence that the stages are more nearly equal in those regions.

The stress test still shows where the classification is most sensitive to the reference set. In the eastern region, where the reference set is largest, less than a third of observations lie on the regional frontier, the two frontiers agree in 90.0% of observations, and the weaker-Stage-C share moves only from 82.5% to 72.5%. The median ratio $\Gamma_{o,F}$ is 1.000 in the east, 0.82 in the central region and 0.78 in the west, so national-frontier Stage F efficiencies in the central and western groups lie appreciably below what the same provinces score against their own peers. Under *S*00, all seven observations that the national frontier classifies as a weaker Stage F belong to Chongqing and Sichuan, and six of them become stage equality within the western region; four further observations are classified as Stage F weaker only under *S*11, namely, Chongqing in 2019, Jiangxi in 2018, Ningxia in 2020 and Hubei in 2021. This shift should be read with the same caution, because 41.7% of western observations lie on the regional frontier.

5.7. Sensitivity to Additional Design Variations

Table 12 consolidates the additional variations. Moderate changes in the stage weights leave scores and rankings close to the baseline: the MAD is 0.0242, the minimum rank correlation is 0.9885, no province moves more than three positions, and no classification changes.

Table 12. Additional sensitivity analyses of system efficiency.

Design Choice	Pairs	Mean MAD	Largest Single Difference	Min. Annual ρ_s	Max. Rank Change	Classification Change
Stage weights 0.4/0.6 and 0.6/0.4	448	0.0242	0.0557	0.9885	3	0.0%
Monetary output: contract value replacing transfer income	224	0.1112	0.7165	0.6888	22	10.3%
Reference set: yearly available	224	0.0222	0.8793	0.8386	22	0.9%
Reference set: remove Tibet from the expanded 31-province reference set	90	0.0366	0.9696	0.8109	28	3.3%
Timing *S*10 versus *S*00	112	0.0970	0.4901	0.7702	16	3.6%
Intermediate link: two-year patent pool, 2019–2021	84	0.0171	0.1107	0.9868	4	0.0%
Intermediate link: three-year patent pool, 2019–2021	84	0.0207	0.1293	0.9813	5	0.0%
Intermediate link: two-year patent pool, 2020–2021	56	0.0178	0.1107	0.9868	4	0.0%
Intermediate link: three-year patent pool, 2020–2021	56	0.0252	0.1293	0.9813	5	0.0%

The remaining dimensions are more consequential. Replacing realized transfer income with contract value yields a MAD of 0.1112, a minimum rank correlation of 0.6888, a movement of one observation by 0.7165, rank movements of up to 22 positions, and a 10.3% change in weaker stage classifications. Contract value records the amounts agreed in signed transactions, whereas realized income records receipts actually booked, so the two indicators measure different stages of transaction completion and are not interchangeable. Extending the reference set to every admissible province changes average scores only slightly (MAD 0.0222) but moves individual province-year observations by up to 0.8793 and individual provinces by up to 22 positions. Removing Tibet from the extended reference set yields a MAD of 0.0366 but moves one observation by 0.9696 and one province by 28 positions. That observation is Qinghai in 2020, whose system efficiency rises from 0.0304 to 1.000 when Tibet is removed: every benchmark that improves on Qinghai under VRS draws on Tibet, so without Tibet, Qinghai is efficient. Frontier composition therefore affects individual benchmarks even when aggregate statistics appear stable, and under VRS, a single extreme unit can determine whether a small unit is efficient.

Replacing the annual patent flow with a rolling pool of grants preserves the stage diagnosis. On the 2019–2021 window, a two-year pool produces a MAD of 0.0171, a minimum annual rank correlation of 0.9868 and a largest single difference of 0.1107, and a three-year pool 0.0207, 0.9813 and 0.1293; the largest rank changes are four and five positions, no weaker stage classification changes, and the share of observations in which commercialization is the weaker stage remains 83.3%. In the 2020–2021 window, the MADs are 0.0178 and 0.0252; there are no classification changes, and the share remains 82.1%. Stage C's weaker position therefore persists under the two tested rolling-pool definitions.

5.8. Application-Level Outputs: Rank Volatility and Frontier Projections

Two application-level outputs follow from these results: provincial rankings and frontier projections. Figure 7 shows provincial ranks and mean system efficiency under the two alignments, and Table A1 in Appendix A reports the province-level values. Under *S*00, Jiangsu is the only province on the frontier over the four-year average, and the ratio between the highest and lowest provincial averages exceeds 25. At the province-year level, 33 of the

112 matched observations change rank by at least five positions between *S*00 and *S*11 and six by at least ten, with a maximum change of 15; under *S*11, Beijing joins Jiangsu on the frontier. In 25 of the 28 provinces, mean Stage C efficiency is below mean Stage F efficiency under *S*00; in one, they are equal; and in Chongqing and Sichuan, Stage C is higher, so a uniform characterization of weak commercialization is not supported at the provincial level, although it describes the sample average.

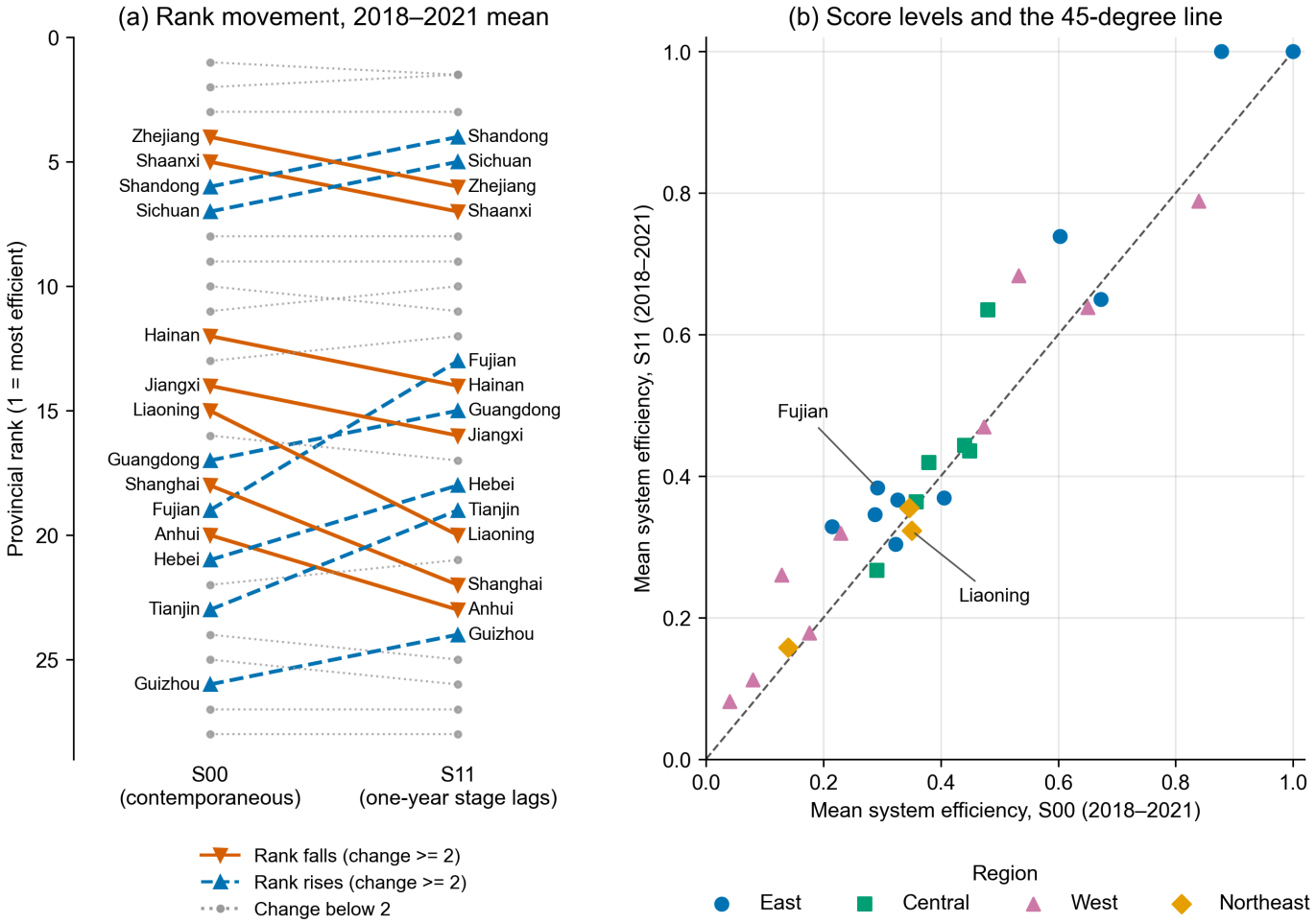


Figure 7. Provincial ranks and mean system efficiency under the two alignments, 2018–2021: (**a**) rank movement, with provinces whose rank changes by at least two positions highlighted; (**b**) mean system efficiency under *S*11 against *S*00, with the 45-degree line.

Table 13 reports aggregate observed quantities and the system-benchmark projections. Over 2018–2023, the 28 provinces spent RMB 703.1 billion on university R&D projects at 2018 prices, were granted 789,462 invention patents, and recorded RMB 25.9 billion of realized technology-transfer income through 135,484 contracts [53]. Realized income thus amounts to 3.7% of R&D project expenditure. Over 2018–2021 under *S*00, realized technology-transfer income of RMB 14.84 billion compares with a system-benchmark projection of RMB 31.84 billion, a model-dependent projection gap of RMB 17.00 billion (53.4%); the corresponding contract gap is 98.9 thousand contracts (59.1%). Relative to university R&D project expenditure of RMB 432.26 billion, realized income is 3.43%, and the projection gap is 3.93%. The gap is right-skewed across units: five provinces account for 36.0% of the total, and the median province-specific gap is RMB 0.58 billion (Appendix A).

Table 13. Aggregate observed quantities and system-benchmark projections of the provincial university research system (2018 prices).

Quantity	Unit	*S*00, 2018–2021	*S*11, 2018–2021	*S*00, 2018–2023
Realized transfer income	RMB billion	14.84	19.66	25.86
Benchmark target income	RMB billion	31.84	40.56	55.20
Benchmark income shortfall	RMB billion	17.00	20.89	29.34
Shortfall share of target	%	53.4	51.5	53.1
Technology-transfer contracts	thousand	68.4	110.4	135.5
Benchmark target contracts	thousand	167.3	238.8	294.5
University R&D expenditure	RMB billion	432.26	432.26	703.08
Realized income/R&D	%	3.43	4.55	3.68
Benchmark shortfall/R&D	%	3.93	4.83	4.17

The projection depends strongly on the benchmark constraints. Under *S*11, whose outcomes are dated two years later and whose realized income is correspondingly larger (RMB 19.66 billion), the gap is RMB 20.89 billion; as a share of the benchmark target, it is 51.5%, against 53.4% under *S*00. More importantly, the system projection is measured at the benchmark link flow (Remark 4): because the benchmark link flow exceeds the observed patent count in 91 of the 112 observations, the projection need not be attainable with the observed annual patent flow. Constraining the benchmark yields three projections from three different programs: the linked-network projection gives RMB 17.00 billion; the first conditional program of Section 3.6, which bounds the weighted benchmark link flow by the observed patent count and evaluates only the commercialization stage on the full yearly reference set, gives RMB 11.47 billion; and the second, which admits only reference units with no more observed patents than the evaluated province, gives RMB 5.86 billion. Relative to university R&D project expenditure, they range from 1.36% to 3.93%. The conditional programs support neither a Stage F efficiency nor a stage classification, and the restricted-peer projection is conditional on its comparison set. The network projection is unique in exact arithmetic (Section 5.4), and for the two conditional programs, the numerical bounds over their optimal sets differ by at most 1.8×10^{-8} thousand RMB per observation in absolute value, so none of the three amounts depends on which optimum the solver returns.

6. Discussion

The results separate into three layers. The models identify key features: the identity of the Stage F efficiencies under BCC and SBM, the signs of two specification steps, and the composition of the network classification under link expansion. Other features are robust empirical findings: the predominant classification of commercialization as the weaker stage holds under the three production specifications for *S*00 and *S*11 and, for the network SBM, across the nine-cell lag grid, and the stage decomposition is proved unique, in exact arithmetic, for all 280 principal solutions. A third layer is specification-dependent: scores, rankings, classification frequency and the size of the frontier projection all vary with the production representation, the alignment, the benchmark link flow and the reference set. The findings depend on the maintained production structure, in particular, on granted invention patents as the intermediate link and on a single national frontier shared by provinces whose university systems and local industrial demand differ substantially.

6.1. Where Specification Dependence Concentrates, and Why

Specification sensitivity concentrates in Stage C for structural reasons the models make explicit. Stage F has a single output, so its BCC and SBM scores coincide (Remark 1), and the network program can only raise its score (Remark 3(b)), here by 0.005–0.012 on average. Stage

C has two outputs whose shortfalls need not be proportional, so the slacks-based measure can only lower its score relative to BCC (Remark 3(a)), here by about 0.11 on average and by 0.314 and 0.246 for Gansu and Hebei, whose contract and income shortfalls diverge. In the network program, Stage C is in addition evaluated at the benchmark link flow (Remark 4), which in this sample exceeds the observed patent count for most observations; the move to the network program, which combines this benchmark link flow with joint system optimization, lowers mean Stage C efficiency by another 0.10–0.11. Because the system efficiency is bounded by the weaker stage efficiency and the weighted arithmetic mean (Remark 2(b)), the concentration of sensitivity in Stage C propagates into system efficiency and provincial rankings.

Two consequences follow. First, the classification "commercialization is weaker" carries two readings. Conditional on the observed patent count, it holds for 71–73% of observations; at the benchmark link flow, for 83–86%. The latter reading partly reflects a structural feature of the as-output network model: every unit whose link is not expanded is classified as Stage C weaker or equal, and the report should state which reading it refers to. Second, the radial choice carries substantive weight, while the linked specification provides one coherent representation when the transfer process is treated as a whole and stage-wise specifications remain informative when a single stage is at issue. Divisional efficiencies and projections in network models depend on the formulation [16,17,28]. A reported estimate should therefore state its specification. The predominance of weaker commercialization agrees with earlier evidence that downstream transformation lags upstream research in Chinese universities [8,9,43]. The present results add that the measured size of that lag depends on the representation and the benchmark link flow.

6.2. Decomposition Non-Uniqueness: Why It Does Not Bind Here

Proposition 1 guarantees that the two auxiliary programs return exact bounds, and Remark 5 states when they are non-trivial: the benchmark link flow must vary along a direction of the feasible set in which the gain in one stage expansion is exactly offset, at the weight ratio, by the loss in the other. Example 1 shows that this can happen and that the procedure then reports an indeterminate classification. In the provincial data, the interval widths lie below the feasibility tolerance in all 280 principal solutions, and the exact certificates prove both the stage decomposition and the income projection unique. Non-uniqueness of the stage decomposition is thus a theoretical possibility that does not occur in this sample. The procedure remains useful because it turns the uniqueness of the reported decomposition from an assumption into a proven property, and it would flag indeterminate units in data with tied or discretized values.

6.3. Temporal Alignment and Window Composition

Research inputs, patent grants, contract signatures and realized receipts occur at intervals that annual provincial data do not reveal. The lag grid shows that alignment shifts scores by roughly 0.09–0.12 MAD and rankings by up to 16 positions, without a monotone effect of span and without overturning the predominant classification. The differences between patent-side and commercialization-side placements of the same lag are too small to support a mechanism-specific interpretation. Nor does the positive average difference between sequential and contemporaneous alignment show that lags raise efficiency, because different alignments compare provinces against differently composed annual frontiers. The consequence is procedural: report any alignment with its window.

6.4. Sensitivity to the Reference Set

Under the national frontier and *S*00 on the 2018–2021 window, the seven province-year observations with a weaker Stage F all lie in the western region and amount to 19.4% of

that region's 36 observations, or 6.3% of the sample; six of the seven become stage equality against their regional peers, so this subset is identified only under the national benchmarking choice. The ordering of regions is consistent with spatial and regional evidence on Chinese university and regional innovation systems [39,49,62].

Two boundaries apply. The first is the sample: three western provinces are excluded from the fixed panel because they do not record strictly positive values in every year, and their exclusion is not random with respect to transfer activity, so the western averages describe nine of its twelve provinces and the national figures describe provinces with continuous transfer activity. The second is the frontier: one national frontier treats twenty-eight provinces with very different university systems and local industrial demand as facing the same attainable technology, so part of what is scored as inefficiency may reflect environment rather than management. Re-benchmarking within regions sharply reduces the weaker-Stage-C share, especially in the smallest groups, and, because the equality class coincides with the on-frontier class, it cannot distinguish a genuine technology gap from the loss of discrimination that a small peer set brings. A common rule of thumb requires the number of units to be at least $2mp$ for m inputs and p outputs [19]. For the two inputs and two final outputs of the system, this threshold is 8, so the central (6) and northeastern (3) groups fall short, and the eastern (10) and western (9) groups only exceed it; the rule is a heuristic, and the high on-frontier shares are the more direct evidence of lost discrimination. Modest perturbations of the national reference set have small average but large individual effects, as the Qinghai case shows.

6.5. Implications for Benchmarking Practice

The evidence supports a procedural recommendation. When DEA rankings are used as benchmarks, the production specification, the benchmark link flow at which a downstream stage is evaluated, the temporal alignment, and the reference set should be reported alongside the scores, because together they move Stage C enough to reorder provinces, and units whose positions are unstable across defensible specifications warrant case-level review before any decision follows from the ranking. Output definitions matter as well: contract value and realized income capture different stages of transaction completion and are not substitutes. Report frontier projections as model-dependent projection gaps: the system-optimal projection, the projection with the benchmark link flow bounded by the observed patent count, and the projection with a restricted peer set answer different questions and differ by about a factor of three, and none establishes an amount recoverable from existing patents. A low Stage C efficiency locates where the measured shortfall lies under the maintained frontier, not which intervention would reduce it.

6.6. Limitations

The study has seven limitations. First, every stage comparison is output-oriented. No input-oriented network model is estimated, so we do not assess the robustness of the stage classifications to orientation; doing so would require a model that treats the intermediate flow as an input link with its own slack. Second, the reference-set comparison is a stress test, not a decomposition. The regional frontiers are estimated from three to ten provinces, leaving between 27.5% and 66.7% of their own observations on the frontier and therefore discriminate less than the national frontier; separating a genuine technology gap from the discrimination loss of a small peer group would require group-frontier or meta-frontier estimation on larger groups than these data support. Third, the panel covers six years, so specifications with a three-year total span leave only three common starting years and cannot capture project-level gestation periods; longer panels or patent-level linkage would allow heterogeneous lag structures to be estimated rather than assumed. Fourth, the analysis is conducted at the provincial level and conceals heterogeneity among universities. Fifth, granted invention

patents are the only intermediate variable and an imperfect proxy for transferable technology: a patent is neither necessary nor sufficient for transfer, provincial patent grants and transfer receipts do not refer to the same set of transferred projects, and the counting rule is sensitive to examination policy. Sixth, the two stage measures are constructed differently: the Stage F measure is a single output ratio, whereas the Stage C measure averages two output ratios at the benchmark link flow, and ratios with small denominators can be large (under $S00$, the median Stage C expansion is 3.9, the largest is 74.8, and 18 of the 112 observations exceed 10). Section 5.5 shows that the predominant classification survives ratio-dominance checks, the removal of small-denominator observations and single-output re-estimation. However, about one in five "Stage C weaker" labels depends on the equal weighting of contracts and income, and the frequency of the label varies by up to 5.4 percentage points relative to the baseline on the same observations across these checks. Seventh, the analysis is descriptive and benchmarking-oriented: it locates the concentration of inefficiency and the size of the frontier projection, but not their causal determinants. The optimal-face intervals are algebraic ranges compatible with the system optimum, not sampling confidence intervals; with 28 DMUs per annual frontier, formal statistical inference on efficiency levels would require additional assumptions on the data-generating process and an appropriate inferential procedure, such as bootstrap methods for nonparametric frontier estimators [36,63].

7. Conclusions

This study examined how the stage diagnosis of a two-stage network efficiency model depends on the specification of the underlying optimization problem, holding the data fixed for 28 Chinese provinces.

For the output-oriented network SBM, fixing the system value at its optimum makes the commercialization expansion an affine function of the formation expansion on the optimal face. Two auxiliary linear programs therefore bound both stage efficiencies and their gap exactly, and the decomposition can be non-unique only if the benchmark link flow varies over the optimal face. These properties rest on the additive aggregation of this model and need not carry over. For all 280 principal solutions, exact rational-arithmetic certificates prove the decomposition unique, so the instability observed in this sample arises from the maintained specifications rather than from the optimizer.

The mean absolute change in Stage C efficiency is about 0.11 at each production-specification step, whereas that in Stage F efficiency is at most 0.012. Commercialization is the predominant weaker stage under all three production specifications for the principal alignments and across the network lag grid, but its share rises from 63–65% under the radial model to 83–86% under the network SBM, partly for structural reasons, and falls to 55.4% under regional frontiers, where it cannot be separated from the reduced discrimination of small peer sets. Efficiency benchmarks should therefore report the production specification, benchmark link flow, temporal alignment and reference set together with each score. Extending the face bounds to input-oriented network models and estimating heterogeneous lags from project-level data are natural next steps.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/math14214085/s1, Data S1: compiled province-year panel, CPI deflators, aligned model inputs, unit-level results of the principal, lag-grid and regional analyses, and the exact rational-arithmetic certificates of the 280 principal network SBM solutions.

Author Contributions: Conceptualization, G.H.; methodology, G.H.; software, G.H.; validation, G.H. and J.H. (Jie Huang); formal analysis, J.H. (Jie Huang); investigation, G.H. and J.H. (Jie Huang); resources, J.H. (Jie Huang); data curation, C.S. and J.H. (Jingli Huang); writing—original draft preparation, G.H. and J.H. (Jie Huang); writing—review and editing, G.H., C.S. and J.H. (Jie Huang); visualization,

G.H.; supervision, J.H. (Jie Huang); project administration, J.H. (Jie Huang); funding acquisition, J.H. (Jie Huang). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Humanities and Social Sciences Research Project of the Ministry of Education of China under grant 22YJCZH061.

Data Availability Statement: The raw data analysed in this study are publicly available in the annual Compilation of Science and Technology Statistics of Higher Education Institutions published by the Ministry of Education of China at http://www.moe.gov.cn/jyb_xwfb/moe_2082/siju/kxjsxxs/gdxxkjtjzl/ (accessed on 3 November 2026); the consumer price indices used for deflation are available from the National Bureau of Statistics of China at https://www.stats.gov.cn/. The compiled panel, the processed model inputs, the unit-level results and the exact rational-arithmetic certificates supporting the conclusions of this article are provided in Supplementary Data S1. The computation code is available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Province-Level Results

Table A1 reports province-level network SBM results for the common 2018–2021 window under the two principal alignments, as summarized in Section 5.8. Provinces whose mean rank changes by at least two positions are bolded; no other selection is applied. Ranks are descending average ranks (mid-ranks for ties) within each specification, so the two frontier provinces under *S*11 share rank 1.5, and a positive rank change denotes a lower rank under *S*11.

Table A1. Provincial network SBM results under the two temporal alignments, 2018–2021 means.

Province		Stage Means (***S*00**)		System Efficiency		Provincial Rank		ΔRank
Name	**Region**	**Stage F**	**Stage C**	***S*00**	***S*11**	***S*00**	***S*11**	(***S*11 − *S*00**)
Jiangsu	East	1.0000	1.0000	1.0000	1.0000	1	1.5	+0.5
Beijing	East	0.9439	0.8450	0.8775	1.0000	2	1.5	−0.5
Chongqing	West	0.7862	0.9110	0.8394	0.7887	3	3	0
Zhejiang	East	0.9492	0.5223	0.6724	0.6498	4	6	+2
Shaanxi	West	0.8203	0.5393	0.6502	0.6387	5	7	+2
Shandong	East	0.9348	0.4560	0.6031	0.7388	6	4	−2
Sichuan	West	0.4847	0.5927	0.5328	0.6831	7	5	−2
Hubei	Central	0.7115	0.3772	0.4796	0.6350	8	8	0
Ningxia	West	0.5729	0.4433	0.4727	0.4698	9	9	0
Henan	Central	0.9822	0.2970	0.4493	0.4358	10	11	+1
Hunan	Central	0.7477	0.3132	0.4406	0.4435	11	10	−1
Hainan	East	0.9975	0.2625	0.4051	0.3696	12	14	+2
Shanxi	Central	0.9061	0.2497	0.3799	0.4190	13	12	−1
Jiangxi	Central	0.5659	0.2871	0.3583	0.3636	14	16	+2
Liaoning	Northeast	0.5258	0.2636	0.3501	0.3228	15	20	+5
Jilin	Northeast	0.6804	0.2341	0.3456	0.3550	16	17	+1
Guangdong	East	0.4416	0.2589	0.3259	0.3665	17	15	−2
Shanghai	East	0.4165	0.2666	0.3230	0.3040	18	22	+4
Fujian	East	0.6403	0.1973	0.2915	0.3833	19	13	−6
Anhui	Central	0.6103	0.1936	0.2905	0.2670	20	23	+3
Hebei	East	0.7319	0.1822	0.2879	0.3460	21	18	−3
Guangxi	West	0.9350	0.1325	0.2291	0.3195	22	21	−1
Tianjin	East	0.4947	0.1487	0.2146	0.3284	23	19	−4
Gansu	West	0.5182	0.1204	0.1756	0.1787	24	25	+1
Heilongjiang	Northeast	0.4658	0.0831	0.1401	0.1580	25	26	+1
Guizhou	West	0.5614	0.0732	0.1282	0.2604	26	24	−2
Inner Mongolia	West	0.3811	0.0447	0.0798	0.1125	27	27	0
Yunnan	West	0.5570	0.0208	0.0398	0.0814	28	28	0

Note: Provinces in bold are those whose mean rank changes by at least two positions between *S*00 and *S*11.

Under *S*00, Jiangsu (1.000) is followed by Beijing (0.878), Chongqing (0.839), Zhejiang (0.672) and Shaanxi (0.650), while Yunnan (0.040), Inner Mongolia (0.080), Guizhou (0.128), Heilongjiang (0.140) and Gansu (0.176) occupy the lowest five positions. Over the four-year averages, Fujian improves from 19th to 13th, Tianjin from 23rd to 19th and Hebei from 21st to 18th between *S*00 and *S*11, whereas Liaoning falls from 15th to 20th, Shanghai from 18th to 22nd and Anhui from 20th to 23rd; only two provinces move by five positions or more. Chongqing (Stage F 0.786, Stage C 0.911) and Sichuan (Stage F 0.485, Stage C 0.593) are the only two provinces for which mean Stage C efficiency exceeds mean Stage F efficiency under *S*00. This comparison depends on the adopted network specification and frontier and does not directly measure the monetary returns generated by patents.

Figure A1 shows, in panel (a), the twelve provinces with the largest absolute system-benchmark projection gap and, in panel (b), realized income relative to university R&D expenditure for all twenty-eight provinces; the dashed line in panel (b) is the median of that ratio, 2.89%. The two panels differ in province coverage, indicators and horizontal scales. Heilongjiang accounts for RMB 1.67 billion of the projection gap, Anhui for RMB 1.30 billion, Tianjin for RMB 1.20 billion, Fujian for RMB 0.99 billion, and Hubei for RMB 0.96 billion. Relative to research spending, the largest gaps are in Yunnan (17.1% of provincial R&D project expenditure), Guangxi (15.8%), Hebei (15.5%), Shanxi (14.5%) and Gansu (14.4%). These are projections at the system benchmark and carry the same qualifications as the aggregate projections.

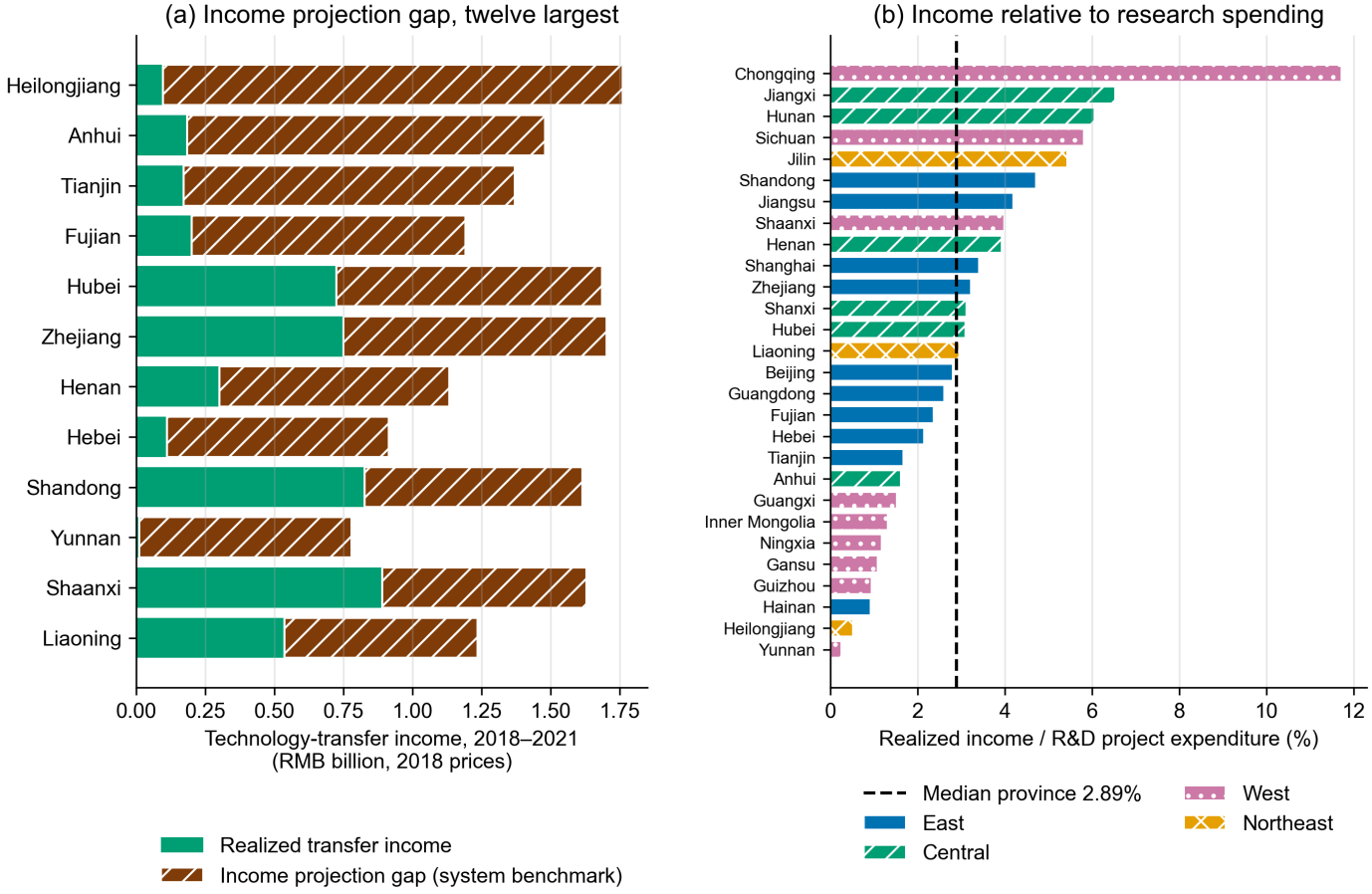


Figure A1. Province-level income projections, 2018–2021: (**a**) system-benchmark income projection gap for the twelve provinces with the largest gaps; (**b**) realized transfer income relative to university R&D project expenditure for all 28 provinces.

Appendix B. Linear-Program Counts

Table A2 lists, by analysis, the 20,002 linear programs covered by the residual and duality-gap checks in Section 4.5. Each network SBM evaluation solves three linear programs (the system program and the two face programs of Equation (14)), and each independent BCC or SBM stage evaluation solves one, that is, four per province-year. Each conditional program is solved together with its two income-bound programs, and the relaxation check solves one system program and eight relaxed face programs per principal solution. Not included are the basis-search programs of the exact certificates, whose results are proved in exact arithmetic, and the programs solved inside DEAPack.

Table A2. Linear programs by analysis and compared pairs.

Analysis	Network Evaluations	Network LPs	Other LPs	Total LPs	Compared Pairs
Principal solutions ($S00$, 2018–2023; $S11$, 2018–2021)	280	840	1120	1960	–
Formation-lag alignment $S10$ (2018–2022)	140	420	560	980	112
Stage weights 0.4/0.6 and 0.6/0.4	448	1344	0	1344	448
Contract value as monetary output	224	672	0	672	224
Extended reference set	238	714	0	714	224
Tibet supplement (with and without Tibet)	183	549	0	549	90
Nominal-data check	280	840	1120	1960	280
Lag grid, nine alignments (Table 8)	728	2184	0	2184	644
Pooled patent link (Table 12)	252	756	0	756	280
National and regional frontiers (Table 11)	224	672	0	672	112
Conditional Stage C programs and income bounds	–	–	672	672	–
Relaxed optimal-face programs ($\delta = 10^{-8}, 10^{-6}$)	–	–	2520	2520	280
Stage-measure checks (Table 10)	833	2499	0	2499	833
Solver cross-check: dual simplex and interior point	560	1680	0	1680	280
DEAPack cross-check (own solutions)	280	840	0	840	280
Total	4670	14,010	5992	20,002	

Notes: In the extended-reference-set analysis, the 238 evaluations comprise the 224 compared evaluations of the 28 fixed provinces (four starting years, two alignments) and 14 evaluations of the added provinces, which are not compared. In the Tibet supplement, 93 evaluations with Tibet (31 provinces, 2020–2022) and 90 without it give 90 compared pairs of the same 30 provinces.

Appendix C. Sample Exclusions and Classification Tolerance

Table A3 lists every non-positive record of the source panel. All of them are zeros of the three commercialization variables in Xinjiang, Qinghai and Tibet; research personnel, R&D expenditure and granted invention patents are strictly positive for all 31 provinces in all years. Table A4 reports the classification shares of Table 9 for alternative values of the tolerance ε in Equation (16).

Table A3. Non-positive records of the source panel and admissible starting years.

Province	Years with $N = I = A = 0$	Admissible Starting Years, $S00$	Admissible Starting Years, $S11$	Extended Reference Set
Xinjiang	2018	2019–2023	2018–2021	2019–2021
Qinghai	2018	2019–2023	2018–2021	2019–2021
Tibet	2018, 2019, 2023	2020–2022	2018–2020	2020

Notes: N, technology-transfer contracts; I, realized transfer income; A, contract value. A starting year is admissible under an alignment when each variable is strictly positive in the year in which it enters (H and R in t, Z in $t + \ell_Z$, N and I in $t + \ell_Z + \ell_C$). The extended reference set admits a province in the starting years 2018–2021 that are admissible under both $S00$ and $S11$.

Table A4. Classification shares (%) under alternative tolerances ε, 2018–2021 (112 observations per row): Stage C weaker/Stage F weaker/equal.

Model, Alignment	Largest $\lvert G_o \rvert$, Equal Class	Smallest Margin, Weaker Classes	$\varepsilon = 0$	$10^{-9} \le \varepsilon \le 10^{-3}$	$\varepsilon = 0.01$	$\varepsilon = 0.05$
Network SBM, *S*00	4.8×10^{-13}	0.038	91.1/7.1/0.9 [a]	85.7/6.3/8.0	85.7/6.3/8.0	85.7/4.5/9.8
Network SBM, *S*11	5.8×10^{-13}	0.0038	85.7/8.9/0.0 [b]	83.0/8.0/8.9	83.0/7.1/9.8	82.1/5.4/12.5
BCC, *S*00	1.1×10^{-15}	0.011	66.1/31.3/2.7	63.4/30.4/6.3	63.4/30.4/6.3	59.8/28.6/11.6
BCC, *S*11	8.0×10^{-15}	0.0030	67.0/31.3/1.8	65.2/29.5/5.4	63.4/28.6/8.0	60.7/25.0/14.3
Independent SBM, *S*00	6.0×10^{-15}	0.013	74.1/25.0/0.9	73.2/20.5/6.3	73.2/20.5/6.3	70.5/16.1/13.4
Independent SBM, *S*11	1.9×10^{-15}	0.0033	72.3/25.0/2.7	71.4/23.2/5.4	70.5/22.3/7.1	69.6/19.6/10.7

Notes: The labels are unchanged for every ε from the largest $|G_o|$ in the equality class up to, but excluding, the smallest margin, $\min G_o^{\min}$ over "Stage C weaker" and $\min(-G_o^{\max})$ over "Stage F weaker". The column $10^{-9} \le \varepsilon \le 10^{-3}$ reproduces Table 9. Across all classifications of the paper, including the lag grid, the regional frontiers, the robustness variations and the checks of Section 5.5, the labels are unchanged for every $\varepsilon \in [1.1 \times 10^{-9}, 1.6 \times 10^{-3}]$. Margins refer to 2018–2021; over all 168 *S*00 solutions (2018–2023), the smallest margin is 0.017. [a] Indeterminate 0.9%. [b] Indeterminate 5.4%.

References

1. Färe, R.; Grosskopf, S. Network DEA. *Socio-Econ. Plan. Sci.* **2000**, *34*, 35–49. [CrossRef]
2. Castelli, L.; Pesenti, R.; Ukovich, W. A classification of DEA models when the internal structure of the decision making units is considered. *Ann. Oper. Res.* **2010**, *173*, 207–235. [CrossRef]
3. Tone, K.; Tsutsui, M. Network DEA: A slacks-based measure approach. *Eur. J. Oper. Res.* **2009**, *197*, 243–252. [CrossRef]
4. Chen, K.; Wupur, A.; Liu, X.; Yang, G.L. Measuring multiple-path technology transfer efficiency in Chinese universities: A network DEA-Tobit approach. *J. Knowl. Econ.* **2025**, *16*, 1464–1493. [CrossRef]
5. Lin, T.Y.; Chiu, S.H.; Yang, H.L. Performance evaluation for regional innovation systems development in China based on the two-stage SBM-DNDEA model. *Socio-Econ. Plan. Sci.* **2022**, *80*, 101148. [CrossRef]
6. Guan, J.; Chen, K. Measuring the innovation production process: A cross-region empirical study of China's high-tech innovations. *Technovation* **2010**, *30*, 348–358. [CrossRef]
7. Yang, G.L.; Fukuyama, H.; Song, Y.Y. Measuring the inefficiency of Chinese research universities based on a two-stage network DEA model. *J. Informetr.* **2018**, *12*, 10–30. [CrossRef]
8. Guo, J.; Liu, J.; Liao, X.; Liu, X. What affects the efficiency of science and technology output and transformation in universities? Evidence from Chinese universities. *IEEE Access* **2023**, *11*, 114201–114219. [CrossRef]
9. Meng, Q.; Li, Y.; Cao, Q. The paradox analysis and functional mechanism between R&D efficiency and transformation effect: Evidence from key universities in China. *Technovation* **2024**, *130*, 102934. [CrossRef]
10. Liu, S.; He, M.; Zhou, Q.; Ma, J.; Zhao, Z. Research on the efficiency of technology transfer in Chinese universities from the perspective of supply and demand. *Socio-Econ. Plan. Sci.* **2026**, *104*, 102382. [CrossRef]
11. Zhang, Q.; Tang, N. The duality of the influence of basic research and applied research on technology transfer performance: Evidence from Chinese universities. *Asian J. Technol. Innov.* **2024**, *32*, 604–631. [CrossRef]
12. Banker, R.D.; Charnes, A.; Cooper, W.W. Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Manag. Sci.* **1984**, *30*, 1078–1092. [CrossRef]
13. Tone, K. A slacks-based measure of efficiency in data envelopment analysis. *Eur. J. Oper. Res.* **2001**, *130*, 498–509. [CrossRef]
14. Ma, D.; Cai, Z.; Zhu, C. Technology transfer efficiency of universities in China: A three-stage framework based on the dynamic network slacks-based measurement model. *Technol. Soc.* **2022**, *70*, 102031. [CrossRef]
15. Lim, D.J.; Kim, M.S. Measuring dynamic efficiency with variable time lag effects. *Omega* **2022**, *108*, 102578. [CrossRef]
16. Kao, C.; Hwang, S.N. Efficiency decomposition in two-stage data envelopment analysis: An application to non-life insurance companies in Taiwan. *Eur. J. Oper. Res.* **2008**, *185*, 418–429. [CrossRef]
17. Liang, L.; Cook, W.D.; Zhu, J. DEA models for two-stage processes: Game approach and efficiency decomposition. *Nav. Res. Logist.* **2008**, *55*, 643–653. [CrossRef]
18. Kao, C. Decomposition of slacks-based efficiency measures in network data envelopment analysis. *Eur. J. Oper. Res.* **2020**, *283*, 588–600. [CrossRef]
19. Dyson, R.G.; Allen, R.; Camanho, A.S.; Podinovski, V.V.; Sarrico, C.S.; Shale, E.A. Pitfalls and protocols in DEA. *Eur. J. Oper. Res.* **2001**, *132*, 245–259. [CrossRef]
20. Sun, Y.; Wang, D.; Yang, F.; Ang, S. Efficiency evaluation of higher education systems in China: A double frontier parallel DEA model. *Comput. Ind. Eng.* **2023**, *176*, 108979. [CrossRef]

21. Chen, Y.; Cook, W.D.; Li, N.; Zhu, J. Additive efficiency decomposition in two-stage DEA. *Eur. J. Oper. Res.* **2009**, *196*, 1170–1176. [CrossRef]
22. Chen, Y.; Li, Y.; Liang, L.; Salo, A.; Wu, H. Frontier projection and efficiency decomposition in two-stage processes with slacks-based measures. *Eur. J. Oper. Res.* **2016**, *250*, 543–554. [CrossRef]
23. Charnes, A.; Cooper, W.W.; Rhodes, E. Measuring the efficiency of decision making units. *Eur. J. Oper. Res.* **1978**, *2*, 429–444. [CrossRef]
24. Tone, K. A slacks-based measure of super-efficiency in data envelopment analysis. *Eur. J. Oper. Res.* **2002**, *143*, 32–41. [CrossRef]
25. Wang, J.; Zhao, M. Evaluation of regional university technology transfer performance in China based on super-efficiency SBM. *Int. J. Technol. Manag.* **2022**, *90*, 54–77. [CrossRef]
26. Chen, X.; Guan, Z.; Yang, G.; Pan, H.; Xiong, X. Evaluation of technology transfer performance for Chinese universities based on a dual-game cross-efficiency model. *Socio-Econ. Plan. Sci.* **2024**, *94*, 101950. [CrossRef]
27. Liu, X.; Wu, X.; Zhang, W. A new DEA model and its application in performance evaluation of scientific research activities in the universities of China's double first-class initiative. *Socio-Econ. Plan. Sci.* **2024**, *92*, 101839. [CrossRef]
28. Chen, Y.; Cook, W.D.; Kao, C.; Zhu, J. Network DEA pitfalls: Divisional efficiency and frontier projection under general network structures. *Eur. J. Oper. Res.* **2013**, *226*, 507–515. [CrossRef]
29. Kao, C. Efficiency decomposition in network data envelopment analysis with slacks-based measures. *Omega* **2014**, *45*, 1–6. [CrossRef]
30. Huang, P.M.; Ma, K.T.; Chien, C.F.; Lee, C.M. Knowledge-based framework for addressing alternative optima in DEA-based project portfolio selection of IC design projects: An empirical study. *Comput. Ind. Eng.* **2026**, *212*, 111637. [CrossRef]
31. Tone, K.; Tsutsui, M. Dynamic DEA: A slacks-based measure approach. *Omega* **2010**, *38*, 145–156. [CrossRef]
32. Tone, K.; Tsutsui, M. Dynamic DEA with network structure: A slacks-based measure approach. *Omega* **2014**, *42*, 124–131. [CrossRef]
33. Li, X.; Hung, H.W.; Lin, I.F.; Lu, C.C. Application of parallel dynamic two-stage directional distance function for R&D efficiency evaluation in China. *Decis. Support Syst.* **2024**, *177*, 114081. [CrossRef]
34. Li, L.B.; Liu, B.L.; Liu, W.L.; Chiu, Y.H. Efficiency evaluation of the regional high-tech industry in China: A new framework based on meta-frontier dynamic DEA analysis. *Socio-Econ. Plan. Sci.* **2017**, *60*, 24–33. [CrossRef]
35. Nasrabadi, N.; Dehnokhalaji, A.; Korhonen, P.; Lokman, B.; Wallenius, J. Robustness of efficiency scores in data envelopment analysis with interval scale data. *Eur. J. Oper. Res.* **2022**, *297*, 1151–1161. [CrossRef]
36. Zhou, L.; Zhang, R.; Li, A. New concepts for bootstrap-based cross-efficiency and relative weight analysis and an application to China's governance–finance–innovation–sustainability system. *J. Clean. Prod.* **2022**, *379*, 134549. [CrossRef]
37. Kadziński, M.; Labijak, A.; Napieraj, M. Integrated framework for robustness analysis using ratio-based efficiency model with application to evaluation of Polish airports. *Omega* **2017**, *67*, 1–18. [CrossRef]
38. Gouveia, M.C.; Henriques, C.O.; Costa, P. Evaluating the efficiency of structural funds: An application in the competitiveness of SMEs across different EU beneficiary regions. *Omega* **2021**, *101*, 102265. [CrossRef]
39. Chen, Z.; Kourtzidis, S.; Tzeremes, P.; Tzeremes, N. A robust network DEA model for sustainability assessment: An application to Chinese provinces. *Oper. Res. Int. J.* **2022**, *22*, 235–262. [CrossRef]
40. Li, G.; Wang, P.; Pal, R. Measuring sustainable technology R&D innovation in China: A unified approach using DEA-SBM and projection analysis. *Expert Syst. Appl.* **2022**, *209*, 118393. [CrossRef]
41. Gong, H.; Nie, L.; Peng, Y.; Peng, S.; Liu, Y. The innovation value chain of patents: Breakthrough in the patent commercialization trap in Chinese universities. *PLoS ONE* **2020**, *15*, e0230805. [CrossRef] [PubMed]
42. Thursby, J.G.; Thursby, M.C. Who is selling the ivory tower? Sources of growth in university licensing. *Manag. Sci.* **2002**, *48*, 90–104. [CrossRef]
43. Cheng, B. Efficiency analysis of 67 Chinese research universities considering inter-university heterogeneity: Evidence from a meta-frontier network SBM DEA model. *PLoS ONE* **2025**, *20*, e0331923. [CrossRef] [PubMed]
44. Chen, T.; Cheng, Y.; Hou, J. The measurement of patent conversion efficiency in China's high-tech industry regions based on a shared input two-stage network DEA model. *Sustainability* **2026**, *18*, 2638. [CrossRef]
45. Bai, X.J.; Li, Z.Y.; Zeng, J. Performance evaluation of China's innovation during the industry–university–research collaboration process—An analysis basis on the dynamic network slacks-based measurement model. *Technol. Soc.* **2020**, *62*, 101310. [CrossRef]
46. Li, X.; Tan, J. Exploring the role of university-run enterprises in technology transfer from Chinese universities. *Manag. Organ. Rev.* **2020**, *16*, 907–943. [CrossRef]
47. Zhang, M.; Zeng, J. Which governance structures are conducive to the performance of TTOs? Evidence from Taiwan. *J. Technol. Transf.* **2025**, *50*, 271–303. [CrossRef]
48. Chen, K.; Wupur, A.; Abudouguli, A.; Yang, G.L. Measuring the knowledge transfer efficiency of social science in Chinese universities from a think tank perspective. *Socio-Econ. Plan. Sci.* **2023**, *90*, 101745. [CrossRef]

49. Fang, H.; Wang, J.; Yang, Q.; Liu, X.; Cao, L. Spatiotemporal characteristics and influencing factor analysis of universities' technology transfer level in China: The perspective of innovation ecosystems. *PLoS ONE* **2022**, *17*, e0270514. [CrossRef] [PubMed]
50. Wang, X. How to promote university technology transfer? A configuration analysis based on technology, organization and environment framework. *PLoS ONE* **2025**, *20*, e0318563. [CrossRef] [PubMed]
51. Huang, Y.; Zhang, J.; Xu, Y. Which configurations promote the high-level university technology transfer? Evidence from China. *High. Educ.* **2026**, *91*, 2119–2142. [CrossRef]
52. Yuan, C.; Li, Y.; Vlas, C.O.; Peng, M.W. Dynamic capabilities, subnational environment, and university technology transfer. *Strateg. Organ.* **2018**, *16*, 35–60. [CrossRef]
53. Ministry of Education of the People's Republic of China. Compilation of Science and Technology Statistics of Higher Education Institutions. 2018–2023 (Annual Volumes). Available online: http://www.moe.gov.cn/jyb_xwfb/moe_2082/siju/kxjsxxs/gdxxkjtjzl/ (accessed on 3 November 2026).
54. Mangasarian, O.L. Uniqueness of solution in linear programming. *Linear Algebra Appl.* **1979**, *25*, 151–162. [CrossRef]
55. National Bureau of Statistics of China. Statistical Communiqués of the People's Republic of China on National Economic and Social Development. 2019–2023. Available online: https://www.stats.gov.cn/ (accessed on 22 September 2026).
56. Spearman, C. The proof and measurement of association between two things. *Am. J. Psychol.* **1904**, *15*, 72–101. [CrossRef] [PubMed]
57. Huangfu, Q.; Hall, J.A.J. Parallelizing the dual revised simplex method. *Math. Program. Comput.* **2018**, *10*, 119–142. [CrossRef]
58. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array programming with NumPy. *Nature* **2020**, *585*, 357–362. [CrossRef] [PubMed]
59. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [CrossRef] [PubMed]
60. McKinney, W. Data structures for statistical computing in Python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010; pp. 56–61. [CrossRef]
61. Wang, D. DEAPack: A Comprehensive Python Toolkit for DEA-Based Efficiency and Productivity Analysis. Software, Version 2.0.1, 2026. Released 10 August 2026. Available online: https://pypi.org/project/DEAPack/2.0.1/ (accessed on 22 September 2026).
62. Xu, K.; Bossink, B.; Chen, Q. Efficiency evaluation of regional sustainable innovation in China: A slack-based measure (SBM) model with undesirable outputs. *Sustainability* **2020**, *12*, 31. [CrossRef]
63. Simar, L.; Wilson, P.W. Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Manag. Sci.* **1998**, *44*, 49–61. [CrossRef]

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