STATISTICAL INFERENCE FOR HICKS–MOORSTEEN PRODUCTIVITY INDICES

Léopold Simar, Valentin Zelenyuk, Shirong Zhao

REPRINT | 2025 / 01

UCLouvain

ISBA | LIDAM Louvain Institute of Data Analysis and Modeling in economics and statistics

ISBA

Voie du Roman Pays 20 - L1.04.01

B-1348 Louvain-la-Neuve

Email : lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/isba/publication

Annals of Operations Research
https://doi.org/10.1007/s10479-024-06288-8

ORIGINAL RESEARCH

Statistical inference for Hicks–Moorsteen productivity indices

Léopold Simar[1] · **Valentin Zelenyuk**[2] · **Shirong Zhao**[3]

Received: 10 December 2023 / Accepted: 10 September 2024

Abstract
The statistical framework for the Malmquist productivity index (MPI) is now well-developed and emphasizes the importance of developing such a framework for its alternatives. In this paper, we try to fill this gap in the literature for another popular measure, known as the Hicks–Moorsteen productivity index (HMPI). Unlike MPI, the HMPI has a total factor productivity interpretation in the sense of measuring productivity as the ratio of aggregated outputs to aggregated inputs and has other useful advantages over MPI. In this work, we develop a novel framework for statistical inference for HMPI in various contexts: when its components are known or when they are replaced with non-parametric envelopment estimators. This will be done for a particular firm's HMPI as well as for the simple mean (unweighted) HMPI and the aggregate (weighted) HMPI. Our results further enrich the recent theoretical developments of nonparametric envelopment estimators for the various efficiency and productivity measures. We also examine the performance of these theoretical results for both the unweighted and weighted mean of HMPI for a finite sample, using Monte-Carlo simulations and also provide an empirical illustration along with the computation code.

Keywords Hicks–Moorsteen productivity index · Data envelopment analysis · Aggregate · Central limit theorem

Mathematics Subject Classification C12 · C14 · C43 · C67

1 Introduction

Two widely used approaches for quantifying changes in productivity over time in the empirical research are the Malmquist productivity index (MPI) (Caves et al., 1982) and the

✉ Valentin Zelenyuk
v.zelenyuk@uq.edu.au

Léopold Simar
leopold.simar@uclouvain.be

Shirong Zhao
shironz@163.com

[1] Institut de Statistique, Biostatistique et Sciences Actuarielles (ISBA), LIDAM, Université Catholique de Louvain, Voie du Roman Pays 20, B1348 Louvain-la-Neuve, Belgium

[2] School of Economics and Centre for Efficiency and Productivity Analysis (CEPA), University of Queensland, Colin Clark Building (39), St Lucia, Brisbane, QLD 4072, Australia

[3] School of Finance, Dongbei University of Finance and Economics, Dalian 116025, Liaoning, China

Published online: 12 October 2024

Hicks–Moorsteen productivity index (HMPI) (Diewert, 1992; Bjurek, 1996). Both MPI and HMPI are commonly estimated using Data Envelopment Analysis (DEA) and Stochastic Frontier Analysis (SFA) estimators, as well as possessing equivalences or approximation relationships with various empirical indices (Fisher, Törnqvist, etc.) under certain conditions.

HMPI has several appealing properties compared to MPI.[1] E.g., HMPI has a total factor productivity (TFP) interpretation (Bjurek, 1996; Grifell-Tatjé & Lovell, 1999) in the sense of measuring productivity as the ratio of aggregated outputs over aggregated inputs, while MPI has TFP properties only under the assumptions of constant returns to scale (CRS) for the technology, but not under variable returns to scale (VRS) (Grifell-Tatjé & Lovell, 1995). It is known that, theoretically, MPI and HMPI coincide under the assumptions of CRS and homotheticity, yet otherwise can produce different results.[2] Another limitation of MPI is that some efficiency components of MPI may not be well defined under VRS-type and FDH-type technologies when working in input or output orientation (Färe et al., 1994a), as is typically done in practice.[3] Also see the work of Briec and Kerstens (2009, 2011) for a more in-depth discussion and Kerstens et al. (2010), Färe et al. (2021) for some empirical examples.[4]

Since its introduction in the 1990s, HMPI and its other variations (e.g., Luenberger-HMPI proposed by Briec & Kerstens, 2004) have been extensively applied to assess productivity changes for a variety of contexts.[5] Recent empirical applications in the literature span various countries, sectors, and industries. Notably, country-level applications are documented in Shen and Valdmanis (2022), Abad and Ravelojaona (2022), Shen et al. (2023) and Deng et al. (2023) while agricultural sector analyses are conducted by O'Donnell (2010), Kerstens and Van de Woestyne (2014), Kerstens et al. (2018), Shen et al. (2019), Baležentis et al. (2021) and Ang et al. (2023). Transportation and logistics applications are explored in See and Li (2015), Seufert et al. (2017) and Abad and Ravelojaona (2017), banking sector evaluations in Epure et al. (2011) and Arjomandi et al. (2014), pharmaceutical industry assessments in Mohammadian and Rezaee (2020), and water company analyses in Sala-Garrido et al. (2018) and Walker et al. (2021). High-tech sector investigations are detailed in Chen et al. (2022) and Chen and Liu (2023), legal sector inquiries in Chen et al. (2024), textile industry studies in Baležentis et al. (2022), education sector examinations in Aparicio et al. (2018), and electricity sector research in Jin et al. (2020). An overall look through the literature suggests that MPI has so far enjoyed a much greater volume of research relative to HMPI. One reason for this, we

[1] It is also worth noting that MPI exhibits its own strengths in comparison to HMPI. Specifically, MPI has various appealing decompositions that dissect productivity into various components, including changes in technical efficiency across variable and constant returns to scale, scale efficiency changes, and technology changes (Grosskopf, 1993; Färe et al., 1994b; Simar & Wilson, 1998, 2019, 2023). Meanwhile, the literature on decomposition of HMPI seems to be still at its early stage (e.g., see Diewert & Fox, 2017).

[2] E.g., refer to the work of Färe et al. (1996) and explore more recent findings in Färe et al. (2021). Therefore, the productivity measured using MPI and HMPI would be different generally and the statistical inference on them could also be different, except under the assumptions of CRS and homotheticity.

[3] As discussed in Simar and Wilson (2019) and Kneip et al. (2021), the utilization of hyperbolic measures helps all components of productivity change, derived through different decompositions of MPI, remain well-defined in wider situations.

[4] We present one example in Section A.1 of the online Appendix A to illustrate this point: when MPI has an infeasible problem but HMPI does not. Moreover, it is also possible that both MPI and HMPI are not well-defined for some points in the technology sets, as illustrated in Section A.2 of the online Appendix A, which appears to be a new insight on HMPI.

[5] It is worth noting that the statistical results developed in this paper for HMPI can be adapted to other variations of HMPI, such as Luenberger-HMPI, which is for a HMPI-type index with directional distance functions instead of the Shephard's distance functions.

think, could be the absence of a rigorous statistical foundation developed specifically for HMPI that would enable reliable statistical inference. Through this study, we aim to address this gap by providing such a theoretical foundation.

Recently, Kneip et al. (2015) laid the groundwork for statistical inferences regarding the simple average of technical efficiency that were estimated via nonparametric envelopment estimators, such as DEA and Free Disposal Hull (FDH) approaches. This enabled many further theoretical developments in efficiency and productivity analysis estimated using nonparametric frontier efficiency methods. Based on Kneip et al. (2015), Kneip et al. (2021) has extensively advanced the statistical inference for DEA estimated MPI, measured with respect to the conical hull of a VRS production frontier. Their contribution encompasses the statistical inference development for both the DEA estimates of a specific firm's MPI and the simple average (unweighted) of DEA estimates of MPI. More recently, Pham et al. (2024) developed analogous framework for the aggregate (weighted harmonic-type mean) of DEA-estimated MPI. However, the framework for statistical inference for HMPI estimates has remained absent.

In this manuscript, building upon prior theoretical contributions, which notably include the works of Kneip et al. (2015, 2021) and Pham et al. (2024), we address a significant gap in the existing literature. Specifically, we delve into establishing the statistical properties of DEA estimators for both the simple mean HMPI and the aggregate HMPI (Mayer & Zelenyuk, 2019) and thus develop the framework for statistical inference for HMPI. We additionally perform numerous Monte-Carlo (MC) experiments to assess the effectiveness of the obtained statistical findings for the simple mean and aggregate HMPI in a wide range of finite samples. Finally, using the most recent Penn World Table data, we examine the productivity changes for countries/regions from 1990 to 2019 to illustrate the usefulness of the developed theoretical results for HMPI in the empirical analyses.

The rest of our study is structured in the following way. In Sect. 2, we provide the theoretical background on Debreu–Farrell distances, individual HMPI, the simple mean and the aggregate HMPI. The estimators are also introduced in this section. Sections 3 and 4 establish the statistical results for the simple mean and aggregate HMPI, respectively. Section 5 performs extensive MC experiments to examine the finite-sample performance of the developed statistical results in Sects. 3 and 4. In Sect. 6, we use one real data set to illustrate the above developed theoretical results. Additional results, including the regularity assumptions for the model, the lemmas used in the main text, all the proofs of the theorems, inference based on geometric means, and additional simulation results, are provided in the Supplementary Document.

2 The theoretical foundation

2.1 The economic model of production

Represent x as a p-dimensional input vector in $\mathbb{R}_+^p$ and y as a q-dimensional output vector in $\mathbb{R}_+^q$. The standard production or technology set can be expressed as

$$\Psi^t = \{(x, y) \mid x \text{ can produce } y \text{ at time } t\}. \tag{2.1}$$

The typical regularity assumptions provided in the online Appendix B are imposed on Ψ^t. The upper boundary of the production set Ψ^t is given as

$$\Psi^{t\partial} := \left\{(x, y) \mid (x, y) \in \Psi^t,\ (x, \lambda y) \notin \Psi^t,\ \forall\, \lambda \in (1, \infty)\right\}, \tag{2.2}$$

which is typically called the *technology frontier*.[6]

The widely used output-oriented Debreu–Farrell distance measure is defined as

$$\lambda(x, y \mid \Psi^t) := \sup\{\lambda > 0 \mid (x, \lambda y) \in \Psi^t\}, \tag{2.3}$$

which provides the maximal feasible proportional increase of all outputs, while keeping the technology and inputs unchanged. The input-oriented Debreu–Farrell distance measure is

$$\theta(x, y \mid \Psi^t) := \inf\left\{\theta > 0 \mid (\theta x, y) \in \Psi^t\right\}, \tag{2.4}$$

which provides the maximal feasible proportional decrease of all inputs, while keeping the technology and outputs unchanged. These measures are reciprocal to the Shephard's output and input distance functions, and hence are primal characterizations of technology, and dual to revenue and cost functions, respectively.[7]

2.2 The Hicks–Moorsteen productivity indices

Let $\mathcal{S}_n = \{(X_i^1, Y_i^1), (X_i^2, Y_i^2)\}_{i=1}^n$ be a randomly selected sample of the input–output pairs for n firms observed in two periods. Further, let $\mathcal{S}_n^t = \{(X_i^t, Y_i^t)\}_{i=1}^n$ be the subsample of $\mathcal{S}_n$ observed only in period t. Each individual firm in period t is assumed to potentially have access to the frontier $\Psi^{t\partial}$. The change in productivity for the firm indexed as i between period 1 to period 2 measured by HMPI (Diewert, 1992; Bjurek, 1996) can be defined as

$$H_i := \left(\frac{\lambda\left(X_i^1, Y_i^2 \mid \Psi^1\right)/\lambda\left(X_i^1, Y_i^1 \mid \Psi^1\right)}{\theta\left(X_i^2, Y_i^1 \mid \Psi^1\right)/\theta\left(X_i^1, Y_i^1 \mid \Psi^1\right)} \times \frac{\lambda\left(X_i^2, Y_i^2 \mid \Psi^2\right)/\lambda\left(X_i^2, Y_i^1 \mid \Psi^2\right)}{\theta\left(X_i^2, Y_i^2 \mid \Psi^2\right)/\theta\left(X_i^1, Y_i^2 \mid \Psi^2\right)}\right)^{-1/2} \tag{2.5}$$

When H_i is less than 1, equal to 1, or greater than 1, it signifies a decrease, no change, or an increase in productivity for firm i over time, respectively. In empirical studies, researchers often focus on the productivity change for the sample over time, which can be measured using the equally weighted geometric averages of the individual HMPI, given by

$$\overline{H}_n := \left(\prod_{i=1}^n H_i\right)^{1/n}, \tag{2.6}$$

where $\overline{H}_n < 1, = 1,$ or > 1, indicates that the productivity for the sample of n firms has decreased, remained unchanged or increased over time.

Now consider the log versions of HMPI. Denote

$$\begin{aligned}\mathcal{H}_i := \log H_i = -\frac{1}{2}\big[&\log\lambda\left(X_i^1, Y_i^2 \mid \Psi^1\right) - \log\lambda\left(X_i^1, Y_i^1 \mid \Psi^1\right)\\ &- \log\theta\left(X_i^2, Y_i^1 \mid \Psi^1\right) + \log\theta\left(X_i^1, Y_i^1 \mid \Psi^1\right)\\ &+ \log\lambda\left(X_i^2, Y_i^2 \mid \Psi^2\right) - \log\lambda\left(X_i^2, Y_i^1 \mid \Psi^2\right)\\ &- \log\theta\left(X_i^2, Y_i^2 \mid \Psi^2\right) + \log\theta\left(X_i^1, Y_i^2 \mid \Psi^2\right)\big],\end{aligned} \tag{2.7}$$

[6] Analogous results can be developed when the true technology is replaced with its conical closures, which we leave to the readers.

[7] All the theories in this paper can be developed in terms of Shephard's distances. Our choice of Debreu–Farrell distances is due to convenience.

for the ith firm, and

$$\overline{\mathcal{H}}_n := \log \overline{H}_n = \frac{1}{n}\sum_{i=1}^{n} \log H_i = \frac{1}{n}\sum_{i=1}^{n} \mathcal{H}_i, \tag{2.8}$$

for the sample of n firms. Clearly $\overline{\mathcal{H}}_n$ is a point estimate of

$$\mu_{\mathcal{H}} := E(\mathcal{H}_i). \tag{2.9}$$

Note that in the above aggregation in (2.8), each observation is treated equally, with weight $1/n$, and so whether it is a very small DMU (firm or country) or a very large one that dominates the economy of the whole group, their productivity score (say 1.1 representing 10% growth) will have exactly the same weight in the aggregate. Thus, such a simple aggregate measure, as an equally-weighted mean of productivity scores, may mis-represent the aggregate productivity of the group due to ignoring the heterogeneity and the economic importance of each DMU being aggregated.

The alternative will be to take the economic importance (e.g., the revenues and/or costs) of each individual into account and consider the aggregate HMPI for the whole sample (Mayer & Zelenyuk, 2019), defined as

$$\begin{aligned}\widetilde{H}_n := \Bigg(& \frac{\sum_{i=1}^{n} S_i^2 \lambda\left(X_i^1, Y_i^2 \mid \Psi^1\right) / \sum_{i=1}^{n} S_i^1 \lambda\left(X_i^1, Y_i^1 \mid \Psi^1\right)}{\sum_{i=1}^{n} W_i^2 \theta\left(X_i^2, Y_i^1 \mid \Psi^1\right) / \sum_{i=1}^{n} W_i^1 \theta\left(X_i^1, Y_i^1 \mid \Psi^1\right)} \\ & \times \frac{\sum_{i=1}^{n} S_i^2 \lambda\left(X_i^2, Y_i^2 \mid \Psi^2\right) / \sum_{i=1}^{n} S_i^1 \lambda\left(X_i^2, Y_i^1 \mid \Psi^2\right)}{\sum_{i=1}^{n} W_i^2 \theta\left(X_i^2, Y_i^2 \mid \Psi^2\right) / \sum_{i=1}^{n} W_i^1 \theta\left(X_i^1, Y_i^2 \mid \Psi^2\right)} \Bigg)^{-1/2},\end{aligned} \tag{2.10}$$

where

$$S_i^t = \frac{p^t Y_i^t}{\sum_{i=1}^{n} p^t Y_i^t}, \tag{2.11}$$

and

$$W_i^t = \frac{w^t X_i^t}{\sum_{i=1}^{n} w^t X_i^t}, \tag{2.12}$$

denote the revenue and cost weights (respectively) for the ith firm at the time t. Additionally, $p^t \in \mathbb{R}_{++}^q$ and $w^t \in \mathbb{R}_{++}^p$ represent the row vector of output and input prices (respectively), assumed to be consistent across different firms at the same time t.

To simplify, we let

$$\begin{aligned} V_{1,i} &= \lambda\left(X_i^1, Y_i^2 \mid \Psi^1\right) p^2 Y_i^2, \quad V_{2,i} = \lambda\left(X_i^1, Y_i^1 \mid \Psi^1\right) p^1 Y_i^1, \\ V_{3,i} &= \theta\left(X_i^2, Y_i^1 \mid \Psi^1\right) w^2 X_i^2, \quad V_{4,i} = \theta\left(X_i^1, Y_i^1 \mid \Psi^1\right) w^1 X_i^1, \\ V_{5,i} &= \lambda\left(X_i^2, Y_i^2 \mid \Psi^2\right) p^2 Y_i^2, \quad V_{6,i} = \lambda\left(X_i^2, Y_i^1 \mid \Psi^2\right) p^1 Y_i^1, \\ V_{7,i} &= \theta\left(X_i^2, Y_i^2 \mid \Psi^2\right) w^2 X_i^2, \quad V_{8,i} = \theta\left(X_i^1, Y_i^2 \mid \Psi^2\right) w^1 X_i^1, \\ V_{9,i} &= p^2 Y_i^2, \; V_{10,i} = p^1 Y_i^1, \quad V_{11,i} = w^2 X_i^2, \quad V_{12,i} = w^1 X_i^1, \end{aligned} \tag{2.13}$$

and denote

$$\mu_s = E(V_{s,i}), \tag{2.14}$$

and

$$\overline{V}_{s,n} = \frac{1}{n}\sum_{i=1}^{n} V_{s,i},\ s = 1, 2, \ldots, 12. \tag{2.15}$$

Now consider the log version of $\widetilde{H}_n$, defined as $\overline{\zeta}_n = \log \widetilde{H}_n$. Analogous to Pham et al. (2024), we can show that

$$\begin{aligned}\overline{\zeta}_n = -\frac{1}{2}\big(&\log \overline{V}_{1,n} - \log \overline{V}_{2,n} - \log \overline{V}_{3,n} + \log \overline{V}_{4,n} \\ &+ \log \overline{V}_{5,n} - \log \overline{V}_{6,n} - \log \overline{V}_{7,n} + \log \overline{V}_{8,n}\big) \\ &+ \log \overline{V}_{9,n} - \log \overline{V}_{10,n} - \log \overline{V}_{11,n} + \log \overline{V}_{12,n},\end{aligned} \tag{2.16}$$

which is a point estimate of

$$\begin{aligned}\zeta = -\frac{1}{2}(&\log \mu_1 - \log \mu_2 - \log \mu_3 + \log \mu_4 \\ &+ \log \mu_5 - \log \mu_6 - \log \mu_7 + \log \mu_8) \\ &+ \log \mu_9 - \log \mu_{10} - \log \mu_{11} + \log \mu_{12}.\end{aligned} \tag{2.17}$$

While the asymptotic theorems presented in this paper bear some analogies to those found in previous literature (Kneip et al., 2021; Pham et al., 2024), our aim is not to merely restate those findings. It is worth noting here that, compared to the simple mean and aggregate MPI (Kneip et al., 2021; Pham et al., 2024), for the simple mean and aggregate HMPI, there is an additional layer of complexity that is also novel. In particular, it is needed to consider the "counterfactual coupling" of inputs observed in one period with outputs observed in a different period, and in particular, reflecting this in the regularity assumptions that make such "counterfactual coupling" well-defined. Furthermore, compared to aggregate MPI in Pham et al. (2024), we have 12 terms instead of 6 terms for the aggregate HMPI, and most of these terms did not appear in Pham et al. (2024). The increased number of terms is due to considering both input orientation and output orientation together in one measure. This additional complexity could be one of the reasons for the absence of rigorous statistical foundation for HMPI, and we aim to fill this gap by developing such a theoretical foundation in this paper.

All the above quantities of $\lambda(X_i^1, Y_i^2 \mid \Psi^1)$, $\lambda(X_i^1, Y_i^1 \mid \Psi^1)$, $\lambda(X_i^2, Y_i^2 \mid \Psi^2)$, $\lambda(X_i^2, Y_i^1 \mid \Psi^2)$, $\theta(X_i^2, Y_i^1 \mid \Psi^1)$, $\theta(X_i^1, Y_i^1 \mid \Psi^1)$, $\theta(X_i^2, Y_i^2 \mid \Psi^2)$, and $\theta(X_i^1, Y_i^2 \mid \Psi^2)$ are the true (or theoretical) quantities, which are not directly observed in empirical analysis and need to be estimated using sample data. In the online Appendix C, we establish the central limit theorems (CLT) results for the case when the various Farrell–Debreu quantities are known. These results will be the benchmark case that will be more generally useful, e.g., when various estimators (e.g., DEA, etc.) are used for estimating the Farrell–Debreu quantities. Also, these results are the baseline results that we rely on to develop the CLT results for DEA estimators discussed in Sects. 3–4.

2.3 DEA estimators of individual and aggregate indices

In the presence of a randomly selected sample $\mathcal{S}_n$, the output-oriented Farrell–Debreu distance $\lambda(x, y \mid \Psi^t)$ and the input-oriented Farrell–Debreu distance $\theta(x, y \mid \Psi^t)$ can be

accomplished using the VRS-DEA estimator, expressed as,

$$\widehat{\lambda}(x, y \mid \mathcal{S}_n^t) = \max_{\lambda, s_1, \ldots, s_n} \Big\{ \lambda \mid \lambda y \leq \sum_{i=1}^{n} s_i Y_i^t,\ x \geq \sum_{i=1}^{n} s_i X_i^t, \sum_{i=1}^{n} s_i = 1,\ \lambda \geq 0,\ \forall\, s_i \geq 0 \Big\}, \tag{2.18}$$

and

$$\widehat{\theta}(x, y \mid \mathcal{S}_n^t) = \min_{\theta, s_1, \ldots, s_n} \Big\{ \theta \mid y \leq \sum_{i=1}^{n} s_i Y_i^t,\ \theta x \geq \sum_{i=1}^{n} s_i X_i^t, \sum_{i=1}^{n} s_i = 1,\ \theta \geq 0,\ \forall\, s_i \geq 0 \Big\}, \tag{2.19}$$

respectively. The corresponding CRS-DEA estimators for $\lambda(x, y \mid \Psi^t)$ and $\theta(x, y \mid \Psi^t)$ can be obtained by deleting $\sum_{i=1}^{n} s_i = 1$ in (2.18) and (2.19), respectively. However, we will focus on VRS-DEA estimators in this study because this is where HMPI has an advantage over MPI.[8]

Plugging the various estimates of Debreu–Farrell distances into the components of $\mathcal{H}_i$ defined in (2.7), we can obtain the individual HMPI estimate as

$$\begin{aligned} \widehat{\mathcal{H}}_i = -\frac{1}{2} \big[& \log \widehat{\lambda}\left(X_i^1, Y_i^2 \mid \mathcal{S}_n^1\right) - \log \widehat{\lambda}\left(X_i^1, Y_i^1 \mid \mathcal{S}_n^1\right) \\ & - \log \widehat{\theta}\left(X_i^2, Y_i^1 \mid \mathcal{S}_n^1\right) + \log \widehat{\theta}\left(X_i^1, Y_i^1 \mid \mathcal{S}_n^1\right) \\ & + \log \widehat{\lambda}\left(X_i^2, Y_i^2 \mid \mathcal{S}_n^2\right) - \log \widehat{\lambda}\left(X_i^2, Y_i^1 \mid \mathcal{S}_n^2\right) \\ & - \log \widehat{\theta}\left(X_i^2, Y_i^2 \mid \mathcal{S}_n^2\right) + \log \widehat{\theta}\left(X_i^1, Y_i^2 \mid \mathcal{S}_n^2\right) \big]. \end{aligned} \tag{2.20}$$

The estimation of the simple mean HMPI, denoted as $\mu_{\mathcal{H}}$, can be conducted through

$$\widehat{\mu}_{\mathcal{H},n} = \frac{1}{n} \sum_{i=1}^{n} \widehat{\mathcal{H}}_i. \tag{2.21}$$

On the other hand, the aggregate HMPI, ζ, can be estimated by

$$\begin{aligned} \widehat{\zeta}_n = -\frac{1}{2} (& \log \widehat{\mu}_{1,n} - \log \widehat{\mu}_{2,n} - \log \widehat{\mu}_{3,n} + \log \widehat{\mu}_{4,n} \\ & + \log \widehat{\mu}_{5,n} - \log \widehat{\mu}_{6,n} - \log \widehat{\mu}_{7,n} + \log \widehat{\mu}_{8,n}) \\ & + \log \widehat{\mu}_{9,n} - \log \widehat{\mu}_{10,n} - \log \widehat{\mu}_{11,n} + \log \widehat{\mu}_{12,n}, \end{aligned} \tag{2.22}$$

[8] The origins of various DEA estimators go back to the works of Farrell (1957), Charnes et al. (1978), Banker et al. (1984), to mention a few. It is one of the most popular approaches, with well-developed statistical properties (as we describe in the next section). We acknowledge that there are also other alternative good estimators in the literature, e.g., see Aigner et al. (1977), Schmidt and Sickles (1984), Daouia and Simar (2007), Amsler et al. (2017), Parmeter and Zelenyuk (2019), Olesen and Ruggiero (2022), Tsionas et al. (2023), to mention just a few. See Ch. 8–16 in Sickles and Zelenyuk (2019) and Kumbhakar et al. (2022a, 2022b) for a more detailed discussion about the various approaches and references. Developing the analogous frameworks for these estimators would require separate papers, though what we present here serves as an important stepping stone for such works.

where

$$\widehat{\mu}_{s,n} = \frac{1}{n}\sum_{i=1}^{n} \widehat{V}_{s,i}, \quad s = 1, 2, \ldots, 8, \tag{2.23}$$

and

$$\widehat{\mu}_{r,n} = \overline{V}_{r,n}, \quad r = 9, 10, 11, 12, \tag{2.24}$$

and where

$$\begin{aligned}
&\widehat{V}_{1,i} = \widehat{\lambda}(X_i^1, Y_i^2 \mid \mathcal{S}_n^1) p^2 Y_i^2, \quad \widehat{V}_{2,i} = \widehat{\lambda}(X_i^1, Y_i^1 \mid \mathcal{S}_n^1) p^1 Y_i^1,\\
&\widehat{V}_{3,i} = \widehat{\theta}(X_i^2, Y_i^1 \mid \mathcal{S}_n^1) w^2 X_i^2, \quad \widehat{V}_{4,i} = \widehat{\theta}(X_i^1, Y_i^1 \mid \mathcal{S}_n^1) w^1 X_i^1,\\
&\widehat{V}_{5,i} = \widehat{\lambda}(X_i^2, Y_i^2 \mid \mathcal{S}_n^2) p^2 Y_i^2, \quad \widehat{V}_{6,i} = \widehat{\lambda}(X_i^2, Y_i^1 \mid \mathcal{S}_n^2) p^1 Y_i^1,\\
&\widehat{V}_{7,i} = \widehat{\theta}(X_i^2, Y_i^2 \mid \mathcal{S}_n^2) w^2 X_i^2, \quad \widehat{V}_{8,i} = \widehat{\theta}(X_i^1, Y_i^2 \mid \mathcal{S}_n^2) w^1 X_i^1,\\
&\overline{V}_{9,n} = \frac{1}{n}\sum_{i=1}^{n} p^2 Y_i^2, \ \overline{V}_{10,n} = \frac{1}{n}\sum_{i=1}^{n} p^1 Y_i^1, \quad \overline{V}_{11,n} = \frac{1}{n}\sum_{i=1}^{n} w^2 X_i^2,\\
&\overline{V}_{12,n} = \frac{1}{n}\sum_{i=1}^{n} w^1 X_i^1.
\end{aligned} \tag{2.25}$$

3 Asymptotic theory for the simple mean HMPI

In principle, the theorems we develop here and in the following section, and the strategy of their proofs, are analogous to those in Kneip et al. (2021) and Pham et al. (2024). However, they require fairly tedious adaptations of the derivations, especially for the aggregation HMPI. and so we leave these to the online Appendix E. Meanwhile, here we will present the essence of the most important results, the spelling out of which is important for practitioners who want to understand the essence of the results and the building blocks needed for the actual computations of the estimates, the bias, the standard errors and the confidence intervals. Further, the statistical inference based on the original, geometric means of the simple mean and aggregate HMPI, are presented in the online Appendix F.

Our first novel result establishes the asymptotic theory for the individual HMPI, which is summarized in Lemma D.3 of the online Appendix D (the analog of Theorem 3.3 in Kneip et al., (Kneip et al., 2021)). The results established in Lemma D.2 of the online Appendix D provide essential tools to derive the statistical properties for the simple mean HMPI, which is provided in Theorem 1 (the analog of Theorem 3.5 in Kneip et al., 2021).

Theorem 1 *Under Assumptions in online Appendix B, as* $n \to \infty$,

$$E(\widehat{\mu}_{\mathcal{H},n}) = \mu_{\mathcal{H}} + C_{\mathcal{H}} n^{-\kappa} + O\left(n^{-\frac{3}{2}\kappa} (\log n)^{\frac{3}{2}\kappa}\right), \tag{3.1}$$

$$\widehat{\mu}_{\mathcal{H},n} - E(\widehat{\mu}_{\mathcal{H},n}) = \overline{\mathcal{H}}_n - \mu_{\mathcal{H}} + o_p(n^{-1/2}), \tag{3.2}$$

$$\widehat{\sigma}^2_{\mathcal{H}} = \frac{1}{n}\sum_{i=1}^{n} (\widehat{\mathcal{H}}_i - \widehat{\mu}_{\mathcal{H},n})^2 \xrightarrow{p} \sigma^2_{\mathcal{H}}. \tag{3.3}$$

The value of κ *is determined by* $2/(p+q+1)$ *when the technology* Ψ^t *displays VRS and* $2/(p+q)$ *when the technology* Ψ^t *exhibits CRS.*

In turn, the results above can be used to establish the CLT for the simple mean HMPI (the analog of Theorem 3.6 in Kneip et al., 2021), which we summarize below.

Theorem 2 *Under Assumptions in online Appendix B, as* $n \to \infty$,

$$\sqrt{n}\left(\widehat{\mu}_{\mathcal{H},n} - \mu_{\mathcal{H}} - C_{\mathcal{H}} n^{-\kappa} - O(n^{-\frac{3}{2}\kappa}(\log n)^{\frac{3}{2}\kappa})\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\mathcal{H}}), \tag{3.4}$$

where recall that $O(n^{-\frac{3}{2}\kappa}(\log n)^{\frac{3}{2}\kappa}) = o(n^{-\kappa})$ *and* $C_{\mathcal{H}}$ *is a constant.*

From (3.1), we can see that $\widehat{\mu}_{\mathcal{H},n}$ is an estimator of $\mu_{\mathcal{H}}$ that exhibits bias, with the bias term in the order of $O(n^{-\kappa})$. The impact of this bias term on the asymptotic behavior of the DEA estimator, $\widehat{\mu}_{\mathcal{H},n}$, can be evaluated by $C_{\mathcal{H}}\sqrt{n}n^{-\kappa}$ in (3.4). When $\kappa > 1/2$, $C_{\mathcal{H}}\sqrt{n}n^{-\kappa}$ goes to zero asymptotically, implying that Theorem 2 can be used directly to make an inference by ignoring the bias term $C_{\mathcal{H}}n^{-\kappa}$; When $\kappa = 1/2$, $C_{\mathcal{H}}\sqrt{n}n^{-\kappa}$ goes to an unknown constant, indicating that Theorem 2 cannot be used directly; When $\kappa < 1/2$, $C_{\mathcal{H}}\sqrt{n}n^{-\kappa}$ goes to infinity asymptotically, implying that Theorem 2 cannot be used directly.

To conduct statistical inference using Theorem 2 for the case $\kappa \leq 1/2$, we adjust the sample size adapting the method of Kneip et al. (2021). Consider $\widehat{\mu}_{\mathcal{H},n_\kappa}$ as a randomly subsampled version (with a sample size of $n_\kappa = \lfloor n^{2\kappa} \rfloor$) of $\widehat{\mu}_{\mathcal{H},n}$.[9] More specifically,

$$\widehat{\mu}_{\mathcal{H},n_\kappa} = \frac{1}{n_\kappa} \sum_{\{i|(X_i^1, Y_i^1, X_i^2, Y_i^2) \in \mathcal{S}_{n_\kappa}\}} \widehat{\mathcal{H}}_i, \tag{3.5}$$

where $\mathcal{S}_{n_\kappa}$ refers to a randomly selected subsample with a sample size of n_κ from the original data set $\mathcal{S}_n$.[10] The statistical properties of $\widehat{\mu}_{\mathcal{H},n_\kappa}$ are established by the subsequent theorem, which is the analog of Theorem B.1 in the work by Kneip et al. (2021).

Theorem 3 *Under Assumptions in online Appendix B, when* $\kappa \leq 1/2$, *as* $n \to \infty$,

$$\sqrt{n_\kappa}\left(\widehat{\mu}_{\mathcal{H},n_\kappa} - \mu_{\mathcal{H}} - C_{\mathcal{H}} n^{-\kappa} - O\left(n^{-\frac{3}{2}\kappa}(\log n)^{\frac{3}{2}\kappa}\right)\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma^2_{\mathcal{H}}\right), \tag{3.6}$$

where $C_{\mathcal{H}}$ *is the same constant as in Theorem 2.*

When $\kappa < 1/2$, the bias term in (3.6), $C_{\mathcal{H}}\sqrt{n_\kappa}n^{-\kappa}$, is stabilized as $n \to \infty$. However, to make an inference, we still need to estimate the bias term $C_{\mathcal{H}}n^{-\kappa}$ for the case $\kappa \leq 1/2$. In a manner analogous to the simple mean MPI context (Kneip et al., 2021), the bias term estimate for $\widehat{\mu}_{\mathcal{H},n}$ can be reliably determined through the generalized jackknife method, which is outlined in the following.

For every $m = 1, 2, \ldots, M$, where $M \ll \binom{n}{\lfloor n/2 \rfloor}$, randomly divide $\mathcal{S}_n$ into two equal-sized subsamples $\mathcal{S}_{n/2,1,m}$ and $\mathcal{S}_{n/2,2,m}$, satisfying $\mathcal{S}_{n/2,1,m} \cap \mathcal{S}_{n/2,2,m} = \emptyset$ and $\mathcal{S}_{n/2,1,m} \cup \mathcal{S}_{n/2,2,m} = \mathcal{S}_n$.[11] Furthermore, for every $l = 1, 2$, separate the subsample $\mathcal{S}_{n/2,l,m}$ by periods 1 and 2 into $\mathcal{S}^1_{n/2,l,m}$ and $\mathcal{S}^2_{n/2,l,m}$ (respectively). For every $l = 1, 2$ and $m = 1, 2, \ldots, M$,

[9] $\lfloor n^{2\kappa} \rfloor$ represents the greatest integer that is no larger than $n^{2\kappa}$. Further $n_\kappa \leq n$ as $\kappa \leq 1/2$.

[10] Please note that in this context, $\widehat{\mathcal{H}}_i$ as defined in (2.20) is calculated individually for each i within the subsample. However, it is computed in relation to all the observations in the sample $\mathcal{S}_n$ instead of the subsample $\mathcal{S}_{n_\kappa}$, and then averaged over all i in that sub-sample (with a size of n_κ) to derive $\widehat{\mu}_{\mathcal{H},n_\kappa}$.

[11] For simplicity, we make the assumption that n is an even number.

calculate

$$\begin{aligned}\widehat{\mu}_{\mathcal{H},n/2,l,m} = \frac{2}{n} \sum_{\substack{\{i|(X_i^1,Y_i^1,X_i^2,Y_i^2)\\ \in \mathcal{S}_{n/2,l,m}\}}} -\frac{1}{2}\Big[&\log \widehat{\lambda}\left(X_i^1, Y_i^2 \mid \mathcal{S}_{n/2,l,m}^1\right) - \log \widehat{\lambda}\left(X_i^1, Y_i^1 \mid \mathcal{S}_{n/2,l,m}^1\right) \\ &- \log \widehat{\theta}\left(X_i^2, Y_i^1 \mid \mathcal{S}_{n/2,l,m}^1\right) + \log \widehat{\theta}\left(X_i^1, Y_i^1 \mid \mathcal{S}_{n/2,l,m}^1\right) \\ &+ \log \widehat{\lambda}\left(X_i^2, Y_i^2 \mid \mathcal{S}_{n/2,l,m}^2\right) - \log \widehat{\lambda}\left(X_i^2, Y_i^1 \mid \mathcal{S}_{n/2,l,m}^2\right) \\ &- \log \widehat{\theta}\left(X_i^2, Y_i^2 \mid \mathcal{S}_{n/2,l,m}^2\right) + \log \widehat{\theta}\left(X_i^1, Y_i^2 \mid \mathcal{S}_{n/2,l,m}^2\right)\Big],\end{aligned} \tag{3.7}$$

and

$$\widehat{\mu}^*_{\mathcal{H},n/2,m} = \frac{1}{2}\left(\widehat{\mu}_{\mathcal{H},n/2,1,m} + \widehat{\mu}_{\mathcal{H},n/2,2,m}\right). \tag{3.8}$$

The bias term estimate for $\widehat{\mu}_{\mathcal{H},n}$ is provided by

$$\widehat{B}_{\mathcal{H},n,\kappa,M} = \frac{1}{M}\sum_{m=1}^{M}(2^{\kappa}-1)^{-1}(\widehat{\mu}^*_{\mathcal{H},n/2,m} - \widehat{\mu}_{\mathcal{H},n}). \tag{3.9}$$

The asymptotic behavior of the bias term estimate, $\widehat{B}_{\mathcal{H},n,\kappa,M}$, is provided by the following theorem (the analog of Eq. (B.6) in Kneip et al., 2021), which can be used to make inferences on the true unobserved simple mean of HMPI.

Theorem 4 *Under Assumptions in online Appendix B, as* $n \to \infty$,

$$\widehat{B}_{\mathcal{H},n,\kappa,M} = C_{\mathcal{H}} n^{-\kappa} + O\left(n^{-\frac{3}{2}\kappa}(\log n)^{\frac{3}{2}\kappa}\right) + o_p(n^{-1/2}), \tag{3.10}$$

where $C_{\mathcal{H}}$ *remains consistent with the constant defined in Theorems 2 and 3.*

Combining Theorems 2, 3 and 4, we can derive the subsequent theorem that can be used to make inferences on the simple mean of HMPI.

Theorem 5 *Under Assumptions in online Appendix B, for* $\kappa \geq 2/5$,

$$\sqrt{n}\left(\widehat{\mu}_{\mathcal{H},n} - \widehat{B}_{\mathcal{H},n,\kappa,M} - \mu_{\mathcal{H}} - R_{\mathcal{H},n,\kappa}\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\mathcal{H}}), \tag{3.11}$$

and if $\kappa < 1/2$, *we have*

$$\sqrt{n_\kappa}\left(\widehat{\mu}_{\mathcal{H},n_\kappa} - \widehat{B}_{\mathcal{H},n,\kappa,M} - \mu_{\mathcal{H}} - R_{\mathcal{H},n,\kappa}\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{\mathcal{H}}), \tag{3.12}$$

where $R_{\mathcal{H},n,\kappa} = O(n^{-\frac{3}{2}\kappa}(\log n)^{\frac{3}{2}\kappa}) = o(n^{-\kappa})$.

Note that $\sigma^2_{\mathcal{H}}$ is not observed, but we can use its empirical version $\widehat{\sigma}^2_{\mathcal{H}}$, as (3.3) shows that $\widehat{\sigma}^2_{\mathcal{H}}$ is a consistent estimate of $\sigma^2_{\mathcal{H}}$. The asymptotic confidence intervals at $100(1-\alpha)\%$ for $\mu_{\mathcal{H}}$ are subsequently expressed as

$$\left[\widehat{\mu}_{\mathcal{H},n} - \widehat{B}_{\mathcal{H},n,\kappa,M} \pm \Phi^{-1}_{1-\alpha/2}\,\widehat{\sigma}_{\mathcal{H}}/\sqrt{n}\right], \tag{3.13}$$

if $\kappa \geq 2/5$ and

$$\left[\widehat{\mu}_{\mathcal{H},n_\kappa} - \widehat{B}_{\mathcal{H},n,\kappa,M} \pm \Phi^{-1}_{1-\alpha/2}\,\widehat{\sigma}_{\mathcal{H}}/\sqrt{n_\kappa}\right], \tag{3.14}$$

if $\kappa < 1/2$.[12] Analogous to Kneip et al. (2021), both (3.13) and (3.14) are applicable when $\kappa = 2/5$, but (3.14) is suggested due to a smaller remainder term, as $\sqrt{n_\kappa} R_{\mathcal{H},n,\kappa}$ converges to zero more quickly than $\sqrt{n} R_{\mathcal{H},n,\kappa}$.

4 Asymptotic theory for the aggregate HMPI

4.1 Basic results

Our first novel result for the aggregate (or weighted mean) HMPI is to derive the statistical properties for the first and second centered moments of $\widehat{V}_{s,i}$ for $i = 1, 2, \ldots, n$ and $s \in \{1, 2, \ldots, 8\}$, which are provided in the Theorem 6 (the analog of Theorem 1 in Pham et al., 2024).

Theorem 6 *Under Assumptions in online Appendix B, for all* $s \in \{1, 2, \ldots, 8\}$, *as* $n \to \infty$, *there exists constants* $0 < C_s < \infty$ *so that for all* $i \in \{1, 2, \ldots, n\}$,

$$E\left(\widehat{V}_{s,i} - V_{s,i}\right) = C_s n^{-\kappa} + O\left(n^{-\frac{3}{2}\kappa} (\log n)^{\frac{3}{2}\kappa}\right), \tag{4.1}$$

and

$$E\left([\widehat{V}_{s,i} - V_{s,i}]^2\right) = o(n^{-\kappa}), \tag{4.2}$$

and for $j \neq i$, $s^* \in \{1, 2, \ldots, 8\}$,

$$\left|E\left([\widehat{V}_{s,i} - E(\widehat{V}_{s,i})] \times [\widehat{V}_{s^*,j} - E(\widehat{V}_{s^*,j})]\right)\right| = o(n^{-1}). \tag{4.3}$$

In turn, from Theorem 6, we can derive the subsequent theorem, which is the analog of Theorem 2 in Pham et al. (2024).

Theorem 7 *Under Assumptions in online Appendix B, for all* $i \in \{1, 2, \ldots, n\}$, $s, s^* \in \{1, 2, \ldots, 8\}$, *and* $r \in \{9, 10, 11, 12\}$, *as* $n \to \infty$,

$$E(\widehat{V}_{s,i}) = \mu_s + C_s n^{-\kappa} + O(n^{-\frac{3}{2}\kappa} (\log n)^{\frac{3}{2}\kappa}), \tag{4.4}$$

$$Cov(\widehat{V}_{s,i}, \widehat{V}_{s^*,i}) = \sigma_{ss^*} + o(n^{-\kappa/2}), \tag{4.5}$$

$$Cov(\widehat{V}_{s,i}, V_{r,i}) = \sigma_{sr} + o(n^{-\kappa/2}). \tag{4.6}$$

Drawing upon Theorem 7, we can deduce the statistical properties for $\widehat{\mu}_s$, and the consistency of $\widehat{\sigma}_{ss^*}$, $\widehat{\sigma}_{sr}$ and $\widehat{\sigma}_{rr^*}$. We summarize this in the next theorem, which is the analog of Theorem 3 in Pham et al. (2024).

[12] The Monte Carlo results provided in Sect. 5 demonstrate that the confidence intervals for the estimated simple mean HMPI tend to under-cover the true values in situations with small sample sizes and large dimensions. This observation aligns with similar patterns observed in unweighted and weighted technical efficiency in the studies of Kneip et al. (2015) and Simar and Zelenyuk (2018), as well as unweighted and weighted MPI in the studies of Kneip et al. (2021) and Pham et al. (2024). To address this issue, one can consider adapting the data sharpening method, as proposed by Nguyen et al. (2022) and Zelenyuk and Zhao (2023), and employing the alternative estimator for the variance, following the approach outlined by Simar et al. (2023, 2024). This adaptation can further enhance the inference of the developed CLT in scenarios characterized by small sample sizes and large dimensions. The same approach can also be extended to improve the inference of the aggregate HMPI presented in Sect. 4.

Theorem 8 *Under Assumptions in online Appendix B, for all* $s, s^* \in \{1, 2, \ldots, 8\}$, $r, r^* \in \{9, 10, 11, 12\}$, *as* $n \to \infty$,

$$E(\widehat{\mu}_{s,n}) = \mu_s + C_s n^{-\kappa} + O\left(n^{-\frac{3}{2}\kappa}(\log n)^{\frac{3}{2}\kappa}\right), \tag{4.7}$$

$$\widehat{\mu}_{s,n} - E(\widehat{\mu}_{s,n}) = \overline{V}_{s,n} - \mu_s + o_p(n^{-1/2}), \tag{4.8}$$

$$\sqrt{n}(\widehat{\mu}_{s,n} - E(\widehat{\mu}_{s,n})) \xrightarrow{d} \mathcal{N}(0, \sigma_{ss}), \tag{4.9}$$

$$\widehat{\sigma}_{ss^*} = \frac{1}{n}\sum_{i=1}^{n}(\widehat{V}_{s,i} - \widehat{\mu}_{s,n})(\widehat{V}_{s^*,i} - \widehat{\mu}_{s^*,n}) \xrightarrow{p} \sigma_{ss^*}, \tag{4.10}$$

$$\widehat{\sigma}_{sr} = \frac{1}{n}\sum_{i=1}^{n}(\widehat{V}_{s,i} - \widehat{\mu}_{s,n})(V_{r,i} - \widehat{\mu}_{r,n}) \xrightarrow{p} \sigma_{sr}, \tag{4.11}$$

$$\widehat{\sigma}_{rr^*} = \frac{1}{n}\sum_{i=1}^{n}(V_{r,i} - \widehat{\mu}_{r,n})(V_{r^*,i} - \widehat{\mu}_{r^*,n}) \xrightarrow{p} \sigma_{rr^*}. \tag{4.12}$$

The results obtained from Theorem 8 yield the following CLT, which allows researchers to permit inference for the aggregate HMPI.

Theorem 9 *Under Assumptions in online Appendix B, as* $n \to \infty$,

$$\sqrt{n}\left(\widehat{\zeta}_n - \zeta - C_\zeta n^{-\kappa} - O\left(n^{-\frac{3}{2}\kappa}(\log n)^{\frac{3}{2}\kappa}\right)\right) \xrightarrow{d} \mathcal{N}(0, \sigma_\zeta^2), \tag{4.13}$$

where recall that $O(n^{-\frac{3}{2}\kappa}(\log n)^{\frac{3}{2}\kappa}) = o(n^{-\kappa})$ *and* C_ζ *is a constant.*

From (9), we can see that $\widehat{\zeta}_n$ is an estimator of ζ with bias on the order of $O(n^{-\kappa})$. The impact of this bias term on the asymptotic behavior of the DEA estimator, $\widehat{\zeta}_n$, can be evaluated by $C_\zeta\sqrt{n}n^{-\kappa}$ in (4.13). When $\kappa > 1/2$, $C_\zeta\sqrt{n}n^{-\kappa}$ goes to zero asymptotically, implying that Theorem 9 can be used directly to make an inference by ignoring the bias term $C_\zeta n^{-\kappa}$. When $\kappa = 1/2$, $C_\zeta\sqrt{n}n^{-\kappa}$ goes to an unknown constant, indicating that Theorem 9 cannot be used directly; When $\kappa < 1/2$, $C_\zeta\sqrt{n}n^{-\kappa}$ goes to infinity asymptotically, implying that Theorem 9 cannot be used directly.

To make an inference using Theorem 9 for the case $\kappa \leq 1/2$, we adjust the sample size adapting the theory from Pham et al. (2024). Consider $\widehat{\zeta}_{n_\kappa}$ as a randomly subsampled version (with a sample size of n_κ) of $\widehat{\zeta}_n$. More specifically,

$$\begin{aligned}\widehat{\zeta}_{n_\kappa} = -\frac{1}{2}(&\log\widehat{\mu}_{1,n_\kappa} - \log\widehat{\mu}_{2,n_\kappa} - \log\widehat{\mu}_{3,n_\kappa} + \log\widehat{\mu}_{4,n_\kappa} \\ &+ \log\widehat{\mu}_{5,n_\kappa} - \log\widehat{\mu}_{6,n_\kappa} - \log\widehat{\mu}_{7,n_\kappa} + \log\widehat{\mu}_{8,n_\kappa}) \\ &+ \log\widehat{\mu}_{9,n_\kappa} - \log\widehat{\mu}_{10,n_\kappa} - \log\widehat{\mu}_{11,n_\kappa} + \log\widehat{\mu}_{12,n_\kappa},\end{aligned} \tag{4.14}$$

where

$$\widehat{\mu}_{s,n_\kappa} = \frac{1}{n_\kappa}\sum_{\{i|(X_i^1,Y_i^1,X_i^2,Y_i^2)\in\mathcal{S}_{n_\kappa}\}}\widehat{V}_{s,i},\ s = 1, 2, \ldots, 8, \tag{4.15}$$

and

$$\widehat{\mu}_{r,n_\kappa} = \frac{1}{n_\kappa}\sum_{\{i|(X_i^1,Y_i^1,X_i^2,Y_i^2)\in\mathcal{S}_{n_\kappa}\}}V_{r,i},\ r = 9, 10, 11, 12, \tag{4.16}$$

and where $\mathcal{S}_{n_\kappa}$ refers to a randomly selected subsample with a sample size of n_κ from the original data set $\mathcal{S}_n$.[13] The statistical properties of $\widehat{\zeta}_{n_\kappa}$ are then established by the theorem below, which is the analog of Theorem 6 in Pham et al. (2024).

Theorem 10 *Under Assumptions in online Appendix B, when $\kappa \leq 1/2$, as $n \to \infty$,*

$$\sqrt{n_\kappa}\left(\widehat{\zeta}_{n_\kappa} - \zeta - C_\zeta n^{-\kappa} - O\left(n^{-\frac{3}{2}\kappa}(\log n)^{\frac{3}{2}\kappa}\right)\right) \xrightarrow{d} \mathcal{N}\left(0, \sigma_\zeta^2\right), \tag{4.17}$$

where C_ζ is the same constant as in Theorem 9.

4.2 Estimating the bias

Drawing inspiration from the approach presented in Pham et al. (2024), the bias of $\widehat{\zeta}_n$ can be reliably determined through the generalized jackknife method, which is outlined in the following.

For every $m = 1, 2, \ldots, M$, where $M \ll \binom{n}{\lfloor n/2 \rfloor}$, randomly divide $\mathcal{S}_n$ into two equal-sized subsamples $\mathcal{S}_{n/2,1,m}$ and $\mathcal{S}_{n/2,2,m}$, satisfying $\mathcal{S}_{n/2,1,m} \cap \mathcal{S}_{n/2,2,m} = \emptyset$ and $\mathcal{S}_{n/2,1,m} \cup \mathcal{S}_{n/2,2,m} = \mathcal{S}_n$.[14] Furthermore, for every $l = 1, 2$, separate the subsample $\mathcal{S}_{n/2,l,m}$ by periods 1 and 2 into $\mathcal{S}^1_{n/2,l,m}$ and $\mathcal{S}^2_{n/2,l,m}$ (respectively). For every $l = 1, 2$ and $m = 1, 2, \ldots, M$, denote $I_i \in \{i | (X_i^1, Y_i^1, X_i^2, Y_i^2) \in \mathcal{S}_{n/2,l,m}\}$ and calculate

$$\begin{aligned}
\widehat{\mu}_{1,l,m} &= \frac{2}{n}\sum_{i\in I_i}\widehat{\lambda}\left(X_i^1, Y_i^2 \mid \mathcal{S}^1_{n/2,l,m}\right)p^2Y_i^2, \quad \widehat{\mu}_{2,l,m} = \frac{2}{n}\sum_{i\in I_i}\widehat{\lambda}\left(X_i^1, Y_i^1 \mid \mathcal{S}^1_{n/2,l,m}\right)p^1Y_i^1,\\
\widehat{\mu}_{3,l,m} &= \frac{2}{n}\sum_{i\in I_i}\widehat{\theta}\left(X_i^2, Y_i^1 \mid \mathcal{S}^1_{n/2,l,m}\right)w^2X_i^2, \quad \widehat{\mu}_{4,l,m} = \frac{2}{n}\sum_{i\in I_i}\widehat{\theta}\left(X_i^1, Y_i^1 \mid \mathcal{S}^1_{n/2,l,m}\right)w^1X_i^1,\\
\widehat{\mu}_{5,l,m} &= \frac{2}{n}\sum_{i\in I_i}\widehat{\lambda}\left(X_i^2, Y_i^2 \mid \mathcal{S}^2_{n/2,l,m}\right)p^2Y_i^2, \quad \widehat{\mu}_{6,l,m} = \frac{2}{n}\sum_{i\in I_i}\widehat{\lambda}\left(X_i^2, Y_i^1 \mid \mathcal{S}^2_{n/2,l,m}\right)p^1Y_i^1,\\
\widehat{\mu}_{7,l,m} &= \frac{2}{n}\sum_{i\in I_i}\widehat{\theta}\left(X_i^2, Y_i^2 \mid \mathcal{S}^2_{n/2,l,m}\right)w^2X_i^2, \quad \widehat{\mu}_{8,l,m} = \frac{2}{n}\sum_{i\in I_i}\widehat{\theta}\left(X_i^1, Y_i^2 \mid \mathcal{S}^2_{n/2,l,m}\right)w^1X_i^1,
\end{aligned} \tag{4.18}$$

and define

$$\begin{aligned}
\widehat{\zeta}_{n/2,l,m} = -\frac{1}{2}\big(&\log\widehat{\mu}_{1,l,m} - \log\widehat{\mu}_{2,l,m} - \log\widehat{\mu}_{3,l,m} + \log\widehat{\mu}_{4,l,m}\\
&+ \log\widehat{\mu}_{5,l,m} - \log\widehat{\mu}_{6,l,m} - \log\widehat{\mu}_{7,l,m} + \log\widehat{\mu}_{8,l,m}\big)\\
&+ \log\widehat{\mu}_9 - \log\widehat{\mu}_{10} - \log\widehat{\mu}_{11} + \log\widehat{\mu}_{12},
\end{aligned} \tag{4.19}$$

and

$$\widehat{\zeta}^*_{n/2,m} = \frac{1}{2}(\widehat{\zeta}_{n/2,1,m} + \widehat{\zeta}_{n/2,2,m}). \tag{4.20}$$

[13] Here, it is important to note that $\widehat{V}_{s,i}$ for $s \in 1, 2, \ldots, 8$ (as expressed in (2.25)), is calculated individually for every firm i within the subsample. However, it is computed with respect to all the data points in $\mathcal{S}_n$ instead of $\mathcal{S}_{n_\kappa}$. Subsequently, it is averaged over all i in that sub-sample (with a size of n_κ) to derive $\widehat{\mu}_{s,n_\kappa}$.

[14] For simplicity, we make the assumption that n is an even number.

The bias term estimate for the aggregate HMPI estimate $\widehat{\zeta}_n$ is provided by

$$\widehat{B}_{\zeta,n,\kappa,M} = \frac{1}{M}\sum_{m=1}^{M}(2^{\kappa}-1)^{-1}\left(\widehat{\zeta}^{*}_{n/2,m} - \widehat{\zeta}_n\right). \tag{4.21}$$

The asymptotic behavior of the bias term estimate, $\widehat{B}_{\zeta,n,\kappa,M}$, is provided by the following theorem.

Theorem 11 *Under Assumptions in online Appendix B, as* $n \to \infty$,

$$\widehat{B}_{\zeta,n,\kappa,M} = C_{\zeta}n^{-\kappa} + O(n^{-\frac{3}{2}\kappa}(\log n)^{\frac{3}{2}\kappa}) + o(n^{-1/2}), \tag{4.22}$$

where C_{ζ} *represents the same constant as specified in Theorems 9 and 10.*

4.3 Making inferences

Combining Theorems 9, 10 and 11, we have the following theorem that can be used to make inferences on the aggregate HMPI.

Theorem 12 *Under Assumptions in online Appendix B, for* $\kappa \geq 2/5$,

$$\sqrt{n}\left(\widehat{\zeta}_n - \widehat{B}_{\zeta,n,\kappa,M} - \zeta - R_{\zeta,n,\kappa}\right) \xrightarrow{d} \mathcal{N}(0,\sigma^2_{\zeta}), \tag{4.23}$$

and if $\kappa < 1/2$, *we have*

$$\sqrt{n_{\kappa}}\left(\widehat{\zeta}_{n_{\kappa}} - \widehat{B}_{\zeta,n,\kappa,M} - \zeta - R_{\zeta,n,\kappa}\right) \xrightarrow{d} \mathcal{N}(0,\sigma^2_{\zeta}), \tag{4.24}$$

where $R_{\zeta,n,\kappa} = O(n^{-\frac{3}{2}\kappa}(\log n)^{\frac{3}{2}\kappa}) = o(n^{-\kappa})$.

However, the true variance σ^2_{ζ} expressed in (C.5) of the online Appendix C is unobserved, but can be estimated using its empirical version, which is denoted as $\widehat{\sigma}^2_{\zeta}$. More specifically,

$$\widehat{\sigma}^2_{\zeta} = [\nabla\widehat{\zeta}_n]'\widehat{\Sigma}[\nabla\widehat{\zeta}_n], \tag{4.25}$$

where $\nabla\widehat{\zeta}_n$ denotes the column vector representing the gradient of $\widehat{\zeta}_n$ concerning $\widehat{\mu}_{s,n}$, i.e., $\nabla\widehat{\zeta}_n = [\frac{\partial\widehat{\zeta}_n}{\partial\widehat{\mu}_{s,n}}]'_{s=1,\ldots,12}$, and where

$$\begin{aligned}
&\frac{\partial\widehat{\zeta}_n}{\partial\widehat{\mu}_{1,n}} = -\frac{1}{2\widehat{\mu}_{1,n}},\ \frac{\partial\widehat{\zeta}_n}{\partial\widehat{\mu}_{2,n}} = \frac{1}{2\widehat{\mu}_{2,n}},\ \frac{\partial\widehat{\zeta}_n}{\partial\widehat{\mu}_{3,n}} = \frac{1}{2\widehat{\mu}_{3,n}},\ \frac{\partial\widehat{\zeta}_n}{\partial\widehat{\mu}_{4,n}} = -\frac{1}{2\widehat{\mu}_{4,n}},\\
&\frac{\partial\widehat{\zeta}_n}{\partial\widehat{\mu}_{5,n}} = -\frac{1}{2\widehat{\mu}_{5,n}},\ \frac{\partial\widehat{\zeta}_n}{\partial\widehat{\mu}_{6,n}} = \frac{1}{2\widehat{\mu}_{6,n}},\ \frac{\partial\widehat{\zeta}_n}{\partial\widehat{\mu}_{7,n}} = \frac{1}{2\widehat{\mu}_{7,n}},\ \frac{\partial\widehat{\zeta}_n}{\partial\widehat{\mu}_{8,n}} = -\frac{1}{2\widehat{\mu}_{8,n}},\\
&\frac{\partial\widehat{\zeta}_n}{\partial\widehat{\mu}_{9,n}} = \frac{1}{\widehat{\mu}_{9,n}},\ \frac{\partial\widehat{\zeta}_n}{\partial\widehat{\mu}_{10,n}} = -\frac{1}{\widehat{\mu}_{10,n}},\ \frac{\partial\widehat{\zeta}_n}{\partial\widehat{\mu}_{11,n}} = -\frac{1}{\widehat{\mu}_{11,n}},\ \frac{\partial\widehat{\zeta}_n}{\partial\widehat{\mu}_{12,n}} = \frac{1}{\widehat{\mu}_{12,n}}.
\end{aligned} \tag{4.26}$$

And, $\widehat{\Sigma}$ is the covariance matrix with the (s,s^*)th element as $\widehat{\Sigma}_{s,s^*} = \widehat{\sigma}_{ss^*}$, where $s, s^* \in \{1,2,\ldots,12\}$.

Drawing upon Theorem 8, we can derive the following theorem which establishes the consistency of the empirical version of the variance of the aggregate HMPI.

Theorem 13 *Under Assumptions in online Appendix B, as* $n \to \infty$,

$$\widehat{\sigma}^2_{\zeta} \xrightarrow{p} \sigma^2_{\zeta}, \tag{4.27}$$

where σ^2_{ζ} *is the true variance of* ζ *and* $\widehat{\sigma}^2_{\zeta}$ *is its empirical version given by (4.25).*

Upon obtaining $\widehat{\sigma}^2_{\zeta}$ from (4.25), based on Theorem 12, the asymptotic confidence intervals at $100(1-\alpha)\%$ for ζ are provided by

$$\left[\widehat{\zeta}_n - \widehat{B}_{\zeta,n,\kappa,M} \pm \Phi^{-1}_{1-\alpha/2}\,\widehat{\sigma}_{\zeta}/\sqrt{n}\right], \tag{4.28}$$

if $\kappa \geq 2/5$ and

$$\left[\widehat{\zeta}_{n_\kappa} - \widehat{B}_{\zeta,n,\kappa,M} \pm \Phi^{-1}_{1-\alpha/2}\,\widehat{\sigma}_{\zeta}/\sqrt{n_\kappa}\right], \tag{4.29}$$

if $\kappa < 1/2$.[15] In line with the methodology proposed by Pham et al. (2024), when $\kappa = 2/5$, both (4.28) and (4.29) are valid. However, it is advisable to use (4.29) due to a smaller remainder term, as $\sqrt{n_\kappa} R_{\zeta,n,\kappa}$ converges to zero more rapidly than $\sqrt{n} R_{\zeta,n,\kappa}$.

5 Monte-Carlo evidence

5.1 Details on MC simulations

The data generation process in our MC simulations is the same as that in Pham et al. (2024). We specify the true technology as

$$y = \psi^1(x) = \prod_{j=1}^{p} (x_1 - 1)^{\beta_j}, \tag{5.30}$$

for the first period and

$$y = \psi^2(x) = (1+\delta) \prod_{j=1}^{p} (x_j - 1)^{\beta_j + \delta}, \tag{5.31}$$

for the second period. Note that the parameter $\delta \geq 0$ is used to control the magnitude of the productivity change. Specifically, the case with $\delta = 0$ indicates no productivity change. The higher the value of δ, the larger increase of the productivity change.

We employ the identical method as Pham et al. (2024) to first generate the time correlated inputs of these two periods, denoted as $x_i^t = (x_{i1}^t, x_{i2}^t, \ldots, x_{ip}^t)$, for $i = 1, \ldots, n$, and $t = 1, 2$, and then generate the time correlated true Debreu–Farrell output-oriented distances, denoted as $\lambda(x_i^t, y_i^t \mid \Psi^t)$ for $t = 1, 2$.[16] For additional information on generating correlated random variables, refer to Appendix EC.3 in Pham et al. (2024). The observed outputs for $t = 1, 2$ can be obtained as

$$y_i^1 = \psi^1(x_i^1)/\lambda(x_i^1, y_i^1 \mid \Psi^1), \tag{5.32}$$

[15] Similar to the findings for the simple mean HMPI, our MC results, available in Sect. 5, reveal that the confidence intervals for the estimated aggregate HMPI exhibit under-coverage of the true values in scenarios with small sample sizes and large dimensions. To enhance the inference of the developed CLT for the aggregate HMPI in such situations, one could consider adapting the data sharpening method, as suggested by Nguyen et al. (2022) and Zelenyuk and Zhao (2023), along with employing the alternative estimator for the variance, following the approach outlined by Simar et al. (2023, 2024).

[16] It is worth noting that it is not necessary that there is a correlation among the true Debreu–Farrell output-oriented distances. Setting the correlation to zero in our simulations confirmed this by leading to the same conclusions. We thank one anonymous referee for bringing up this question.

and

$$y_i^2 = \psi^2(x_i^2)/\lambda(x_i^2, y_i^2 \mid \Psi^2), \tag{5.33}$$

respectively. Thus, a simulated sample can be obtained as $\mathcal{S}_n = \{(x_i^1, y_i^1, x_i^2, y_i^2)\}_{i=1}^n$. Moreover, as the dimension of outputs is 1, we have $\lambda(x_i^1, y_i^2 \mid \Psi^1) = \lambda(x_i^1, y_i^1 \mid \Psi^1) y_i^1/y_i^2$ and $\lambda(x_i^2, y_i^1 \mid \Psi^2) = \lambda(x_i^2, y_i^2 \mid \Psi^2) y_i^2/y_i^1$.

To determine the true values of the simple mean and aggregate HMPI, we need to know the true Debreu–Farrell input-oriented distances, which can be computed using the following methods. Suppose we want to compute the true Debreu–Farrell input-oriented distance of the observation (x, y) toward Ψ^t, i.e., $\theta(x, y \mid \Psi^t)$. Fixing (x, y), $\theta(x, y \mid \Psi^t)$ must be the solution to $f(a) = y - \psi^t(ax) = 0$ with the domain $a \geq a_{min} = \max_{j=1,2,\ldots,p}\{\frac{1}{x_j}\}$. It can be shown that $f(a)$ is a decreasing function, $f(a_{min}) = y > 0$, and $\lim_{a\to+\infty} f(a) = -\infty$. Hence there must be a unique solution to $f(a) = 0$, which can be obtained using the *uniroot* command in *R*. Thus, $\theta(x, y \mid \Psi^t)$ is unique. The true values of the simple mean and aggregate HMPI are obtained by randomly simulating 10, 000, 000 observations and computing it according to (2.8) and (2.16), respectively.

We consider various values of $p \in \{1, 2, 3, 4, 5, 7\}$ and $\delta = \{0.00, 0.02, 0.04\}$. For each value of p, the corresponding values of β_j and input prices w_j for $j = 1, 2, \ldots, p$, and the value of the output price p_1 are presented in Table G.1, where we assume the input and output prices are identical for these two periods. More specifically, we let $w^1 = w^2 = (w_1, \ldots, w_p)$ and $p^1 = p^2 = p_1$. For each set of simulations measured using (n, δ, p), we conduct 1, 000 replications. Moreover, we present the rejection rates and the coverages to evaluate the power and significance of the tests developed in Sects. 3 and 4.

Before showcasing the MC results and to streamline the notation, we have the following notation for the simple mean HMPI.

(i) Using the standard CLTs, i.e., using $\sqrt{n}$ consistency and without the bias correction.
(ii) Using our developed statistical results, i.e., using (3.13) and (3.14) for $p + q < 4$ and $p + q \geq 4$, respectively.
(iii) Using (3.14) for $p + q \geq 4$, but with the recentered version, i.e., using

$$\left[\widehat{\mu}_{\mathcal{H},n} - \widehat{B}_{\mathcal{H},n,\kappa,M} \pm \Phi^{-1}_{1-\alpha/2}\, \widehat{\sigma}_{\mathcal{H}}/\sqrt{n_\kappa}\right]. \tag{5.34}$$

For the aggregate HMPI, we use the following notation,

(i) Using the standard CLTs, i.e., using $\sqrt{n}$ consistency and without the bias correction.
(ii) Using our developed statistical results, i.e., using (4.28) and (4.29) for $p + q < 4$ and $p + q \geq 4$, respectively.
(iii) Using (4.29) for $p + q \geq 4$, but with the recentered version, i.e., using

$$\left[\widehat{\zeta}_n - \widehat{B}_{\zeta,n,\kappa,M} \pm \Phi^{-1}_{1-\alpha/2}\, \widehat{\sigma}_{\zeta}/\sqrt{n_\kappa}\right]. \tag{5.35}$$

5.2 Main results from MC simulations

We present the primary Monte Carlo results for the case $\delta = 0.04$, while the results for $\delta = 0.00$ and $\delta = 0.02$ are reported in the online Appendix G. The results presented in Tables 1, 2, 3 and 4 for the rejection rates and coverages for the simple mean and aggregate HMPI, generally support the developed statistical results in Sects. 3 and 4.

Table 1 Rejection rates for test for the simple mean productivity change $\mu_{\mathcal{H}}$ with $\delta = 0.04$

p	q	n	0.10		0.05		0.01	
			(i)	(ii)	(i)	(ii)	(i)	(ii)
1	1	20	0.544	0.544	0.437	0.437	0.247	0.247
1	1	50	0.809	0.809	0.728	0.728	0.534	0.534
1	1	100	0.936	0.936	0.903	0.903	0.796	0.796
1	1	200	0.997	0.997	0.992	0.992	0.976	0.976
1	1	300	1.000	1.000	1.000	1.000	0.996	0.996
1	1	500	1.000	1.000	1.000	1.000	1.000	1.000
1	1	1000	1.000	1.000	1.000	1.000	1.000	1.000
2	1	20	0.782	0.766	0.704	0.691	0.558	0.541
2	1	50	0.968	0.965	0.944	0.943	0.864	0.860
2	1	100	1.000	1.000	0.999	0.998	0.992	0.991
2	1	200	1.000	1.000	1.000	1.000	1.000	1.000
2	1	300	1.000	1.000	1.000	1.000	1.000	1.000
2	1	500	1.000	1.000	1.000	1.000	1.000	1.000
2	1	1000	1.000	1.000	1.000	1.000	1.000	1.000

p	q	n	0.10			0.05			0.01		
			(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)
3	1	20	0.919	0.752	0.798	0.890	0.675	0.721	0.792	0.504	0.515
3	1	50	0.999	0.966	0.987	0.997	0.940	0.969	0.984	0.858	0.912
3	1	100	1.000	0.998	1.000	1.000	0.995	1.000	1.000	0.980	0.997
3	1	200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000
3	1	300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	1	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	1	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	1	20	0.983	0.839	0.850	0.962	0.757	0.781	0.915	0.583	0.578
4	1	50	1.000	0.974	0.997	1.000	0.957	0.982	1.000	0.851	0.927
4	1	100	1.000	0.996	1.000	1.000	0.993	1.000	1.000	0.974	0.999
4	1	200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000
4	1	300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	1	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	1	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	1	20	0.998	0.850	0.901	0.995	0.786	0.834	0.981	0.623	0.642
5	1	50	1.000	0.980	0.999	1.000	0.963	0.990	1.000	0.891	0.963
5	1	100	1.000	0.996	1.000	1.000	0.992	1.000	1.000	0.975	1.000
5	1	200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	1	300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	1	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	1	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
7	1	20	1.000	0.872	0.949	1.000	0.826	0.890	1.000	0.677	0.699

Table 1 continued

p	q	n	0.10			0.05			0.01		
			(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)
7	1	50	1.000	0.972	1.000	1.000	0.956	1.000	1.000	0.899	0.977
7	1	100	1.000	0.999	1.000	1.000	0.995	1.000	1.000	0.981	1.000
7	1	200	1.000	0.999	1.000	1.000	0.999	1.000	1.000	0.999	1.000
7	1	300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
7	1	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
7	1	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

We first check the power of our developed hypothesis test. For the simple mean HMPI with results presented in Table 1, we find that when $\delta = 0.04$, i.e., when the productivity indeed increases from period 1 to period 2, the rejection rates using our developed statistical results (i.e., using (ii)) increase toward 100% when the sample size n increases, regardless of the dimension p. This result also holds for the aggregate HMPI presented in Table 2. Consequently, the results of the rejection rates presented in Tables 1 and 2 indicate that our developed hypothesis tests in Sects. 3 and 4 are very good at detecting a false null hypothesis.

We then check the significance of our developed hypothesis test. We see that the coverages for the simple mean HMPI presented in Table 3 and the aggregate HMPI presented in Table 4 increase toward the corresponding nominal coverage $(1 - \alpha)$ when the sample size n increases.[17] Moreover, as the sample size n increases, the coverages of the recentered method (iii) are larger than (ii) and increase toward 100% instead of the nominal coverage, even though the width of the estimated confidence intervals using (ii) is equal to that using (iii). The outcome aligns with the Monte Carlo results reported in Kneip et al. (2015) and Pham et al. (2024) for the simple mean technical efficiency and the aggregate MPI (respectively). Furthermore, analogous to Pham et al. (2024), we see that the developed statistical results perform quite well even in small sample sizes for both the simple mean and aggregate HMPI. E.g., when $p = 2$, the nominal coverage is 95% and the sample size $n = 20$, the coverage of the estimated confidence interval is 0.908 for the simple mean HMPI and 0.821 for the aggregate HMPI.

It is worth noting that for both the simple mean and aggregate HMPI, (i) (the standard CLT) has the similar coverage as (ii) for low dimensions (such as $p \leq 4$), but for high dimensions $p \geq 5$, we see that the coverage using (ii) is larger than those using (i) when the sample size $n \geq 300$. This result is consistent with Pham et al. (2024) that the bias of the simple mean and aggregate HMPI might be too small or even canceling out in some special cases, leading to the similar performance of (i) and (ii).

[17] When $p = 3$ and $q = 1$ such that $\kappa = 2/5$, for both the simple mean and aggregate HMPI, we also find that the sub-sampling method (using (3.14) or (4.29)) is better than the full sample method (using (3.13) or (4.28))(this MC result is not presented in the main text), confirming our discussion in Sects. 3 and 4 that the sub-sampling method is better than the full sample method when $\kappa = 2/5$. This result is consistent with Kneip et al. (2015), Simar and Zelenyuk (2018) and Pham et al. (2024) for the simple mean technical efficiency, the aggregate technical efficiency, and the aggregate MPI, respectively.

Table 2 Rejection rates for test for the aggregate productivity change ζ with $\delta = 0.04$

p	q	n	0.10		0.05		0.01	
			(i)	(ii)	(i)	(ii)	(i)	(ii)
1	1	20	0.746	0.746	0.630	0.630	0.420	0.420
1	1	50	0.975	0.975	0.951	0.951	0.853	0.853
1	1	100	1.000	1.000	1.000	1.000	0.990	0.990
1	1	200	1.000	1.000	1.000	1.000	1.000	1.000
1	1	300	1.000	1.000	1.000	1.000	1.000	1.000
1	1	500	1.000	1.000	1.000	1.000	1.000	1.000
1	1	1000	1.000	1.000	1.000	1.000	1.000	1.000
2	1	20	0.923	0.879	0.883	0.823	0.755	0.719
2	1	50	0.997	0.995	0.997	0.995	0.988	0.975
2	1	100	1.000	1.000	1.000	1.000	1.000	1.000
2	1	200	1.000	1.000	1.000	1.000	1.000	1.000
2	1	300	1.000	1.000	1.000	1.000	1.000	1.000
2	1	500	1.000	1.000	1.000	1.000	1.000	1.000
2	1	1000	1.000	1.000	1.000	1.000	1.000	1.000

p	q	n	0.10			0.05			0.01		
			(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)
3	1	20	0.969	0.787	0.812	0.940	0.721	0.740	0.865	0.572	0.591
3	1	50	0.999	0.974	0.982	0.999	0.950	0.973	0.993	0.877	0.923
3	1	100	1.000	1.000	1.000	1.000	0.998	1.000	1.000	0.988	0.999
3	1	200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	1	300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	1	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
3	1	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	1	20	0.998	0.864	0.882	0.993	0.812	0.824	0.974	0.681	0.712
4	1	50	1.000	0.978	0.993	1.000	0.959	0.983	1.000	0.911	0.944
4	1	100	1.000	0.998	1.000	1.000	0.996	1.000	1.000	0.988	1.000
4	1	200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	1	300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	1	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
4	1	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	1	20	1.000	0.868	0.891	0.999	0.816	0.859	0.994	0.691	0.722
5	1	50	1.000	0.984	0.997	1.000	0.970	0.993	1.000	0.929	0.965
5	1	100	1.000	1.000	1.000	1.000	0.997	1.000	1.000	0.990	1.000
5	1	200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.999	1.000
5	1	300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	1	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
5	1	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
7	1	20	1.000	0.896	0.938	1.000	0.842	0.897	1.000	0.705	0.762
7	1	50	1.000	0.987	0.998	1.000	0.979	0.997	1.000	0.937	0.982

Table 2 continued

p	q	n	0.10			0.05			0.01		
			(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)
7	1	100	1.000	0.998	1.000	1.000	0.998	1.000	1.000	0.990	0.999
7	1	200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
7	1	300	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
7	1	500	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
7	1	1000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 3 Coverages of estimated confidence intervals for the simple mean HMPI $\mu_{\mathcal{H}}$ with $\delta = 0.04$

p	q	n	0.90		0.95		0.99b	
			(i)	(ii)	(i)	(ii)	(i)	(ii)
1	1	20	0.898	0.898	0.947	0.947	0.988	0.988
1	1	50	0.883	0.883	0.939	0.939	0.996	0.996
1	1	100	0.890	0.890	0.952	0.952	0.990	0.990
1	1	200	0.883	0.883	0.931	0.931	0.984	0.984
1	1	300	0.897	0.897	0.952	0.952	0.995	0.995
1	1	500	0.897	0.897	0.943	0.943	0.991	0.991
1	1	1000	0.879	0.879	0.936	0.936	0.983	0.983
2	1	20	0.876	0.838	0.944	0.908	0.982	0.968
2	1	50	0.890	0.878	0.950	0.944	0.988	0.984
2	1	100	0.902	0.899	0.953	0.962	0.993	0.991
2	1	200	0.900	0.901	0.949	0.957	0.993	0.991
2	1	300	0.910	0.899	0.953	0.954	0.994	0.990
2	1	500	0.898	0.888	0.956	0.947	0.987	0.989
2	1	1000	0.915	0.911	0.959	0.953	0.994	0.993

p	q	n	0.90			0.95			0.99		
			(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)
3	1	20	0.855	0.806	0.894	0.915	0.884	0.936	0.975	0.961	0.981
3	1	50	0.879	0.865	0.950	0.933	0.922	0.976	0.981	0.978	0.998
3	1	100	0.878	0.901	0.985	0.939	0.956	0.993	0.989	0.986	1.000
3	1	200	0.880	0.909	0.990	0.945	0.956	0.998	0.990	0.990	1.000
3	1	300	0.901	0.902	0.998	0.958	0.956	1.000	0.990	0.994	1.000
3	1	500	0.904	0.911	0.999	0.952	0.956	1.000	0.989	0.993	1.000
3	1	1000	0.888	0.893	0.999	0.941	0.946	1.000	0.986	0.991	1.000
4	1	20	0.896	0.843	0.925	0.940	0.899	0.960	0.976	0.965	0.987
4	1	50	0.881	0.865	0.987	0.931	0.929	0.996	0.993	0.989	1.000
4	1	100	0.886	0.882	0.994	0.941	0.932	1.000	0.985	0.984	1.000
4	1	200	0.885	0.890	1.000	0.943	0.927	1.000	0.993	0.980	1.000
4	1	300	0.910	0.899	1.000	0.950	0.948	1.000	0.992	0.994	1.000
4	1	500	0.877	0.916	1.000	0.932	0.958	1.000	0.983	0.990	1.000
4	1	1000	0.906	0.893	1.000	0.952	0.940	1.000	0.995	0.988	1.000

Table 3 continued

p	q	n	0.90			0.95			0.99		
			(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)
5	1	20	0.881	0.815	0.939	0.941	0.890	0.968	0.984	0.961	0.991
5	1	50	0.872	0.870	0.990	0.937	0.933	0.994	0.988	0.986	0.999
5	1	100	0.910	0.893	0.999	0.956	0.945	1.000	0.992	0.987	1.000
5	1	200	0.906	0.889	1.000	0.950	0.945	1.000	0.988	0.984	1.000
5	1	300	0.874	0.903	1.000	0.934	0.958	1.000	0.986	0.992	1.000
5	1	500	0.894	0.906	1.000	0.949	0.948	1.000	0.987	0.990	1.000
5	1	1000	0.891	0.900	1.000	0.943	0.960	1.000	0.990	0.991	1.000
7	1	20	0.867	0.795	0.941	0.916	0.873	0.967	0.972	0.958	0.989
7	1	50	0.890	0.867	0.998	0.948	0.935	1.000	0.987	0.982	1.000
7	1	100	0.907	0.890	1.000	0.956	0.946	1.000	0.995	0.990	1.000
7	1	200	0.900	0.885	1.000	0.950	0.927	1.000	0.992	0.984	1.000
7	1	300	0.856	0.900	1.000	0.917	0.939	1.000	0.988	0.983	1.000
7	1	500	0.884	0.901	1.000	0.948	0.949	1.000	0.987	0.990	1.000
7	1	1000	0.885	0.896	1.000	0.931	0.940	1.000	0.991	0.984	1.000

6 Empirical illustrations

To conduct a direct comparison with the productivity change measured using MPI as in Zelenyuk and Zhao (2023), in the following illustration we use the same data and conduct the same analyses for HMPI.[18] The commonly employed data set, Penn World Table data (PWT 10.0), is employed to estimate the simple mean and aggregate HMPI. Analogous to Pham et al. (2024) and Zelenyuk and Zhao (2023), we assume that countries/regions produce GDP using labor and capital stock and we estimate separately for all 84 countries (27 developed countries and 57 developing countries) for the period 1990–2019. Analogous to Zelenyuk and Zhao (2023), we conduct the analyses for pairs of years with 5-year intervals and the entire period from 1990 to 2019 to examine the progression of the simple mean and aggregate HMPI. Moreover, when computing aggregate HMPI, we let the revenues on the outputs and expenses on the inputs be equal to the real GDP. Tables 5 and 6 present the estimated results.

First, we can see that for most of the rows, the estimates before the bias correction are largely different from those after the bias correction, illustrating the practical importance in the empirical analyses to correct the bias in estimates for the simple mean and aggregate HMPI. If the bias is not corrected, the productivity change might be over- or under-estimated or even the direction of the productivity change is flipped, leading to biased conclusions. E.g., Table 5 shows that productivity for the whole sample decreased by 5.51% from 2005 to 2010 before the bias correction, which is substantially smaller than 8.90% after the bias correction. One example that demonstrates the flipped direction for the productivity change is that in Table 6 which shows that the productivity for the whole sample increased by 2.34% between 2000 and 2005 before the bias correction, while it decreased by 2.96% after the bias correction.

[18] To provide tools for the practitioners, the code used in this illustration is available in the GitHub. See https://github.com/srzhao89/szz-hm for more details.

Table 4 Coverages of estimated confidence intervals for the aggregate HMPI ζ with $\delta = 0.04$

p	q	n	0.90		0.95		0.99	
			(i)	(ii)	(i)	(ii)	(i)	(ii)
1	1	20	0.890	0.890	0.940	0.940	0.980	0.980
1	1	50	0.872	0.872	0.929	0.929	0.988	0.988
1	1	100	0.905	0.905	0.947	0.947	0.987	0.987
1	1	200	0.889	0.889	0.935	0.935	0.988	0.988
1	1	300	0.913	0.913	0.952	0.952	0.993	0.993
1	1	500	0.888	0.888	0.943	0.943	0.984	0.984
1	1	1000	0.906	0.906	0.940	0.940	0.988	0.988
2	1	20	0.856	0.747	0.922	0.821	0.976	0.915
2	1	50	0.869	0.824	0.926	0.884	0.985	0.957
2	1	100	0.889	0.865	0.942	0.917	0.982	0.978
2	1	200	0.859	0.847	0.921	0.915	0.980	0.980
2	1	300	0.888	0.886	0.943	0.935	0.991	0.985
2	1	500	0.883	0.872	0.937	0.925	0.987	0.988
2	1	1000	0.892	0.881	0.949	0.941	0.990	0.992

p	q	n	0.90			0.95			0.99		
			(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)
3	1	20	0.825	0.733	0.781	0.884	0.817	0.873	0.953	0.919	0.949
3	1	50	0.859	0.813	0.900	0.920	0.893	0.951	0.985	0.967	0.986
3	1	100	0.851	0.862	0.955	0.913	0.909	0.983	0.984	0.978	1.000
3	1	200	0.872	0.883	0.977	0.936	0.937	0.991	0.982	0.983	0.999
3	1	300	0.873	0.875	0.989	0.940	0.931	0.999	0.981	0.991	1.000
3	1	500	0.886	0.887	0.989	0.933	0.948	0.999	0.982	0.988	1.000
3	1	1000	0.862	0.891	0.994	0.921	0.939	1.000	0.981	0.982	1.000
4	1	20	0.838	0.731	0.801	0.905	0.814	0.867	0.965	0.921	0.950
4	1	50	0.825	0.798	0.917	0.908	0.880	0.959	0.973	0.954	0.991
4	1	100	0.848	0.851	0.973	0.910	0.920	0.991	0.974	0.974	0.999
4	1	200	0.868	0.858	0.991	0.928	0.922	0.997	0.983	0.986	1.000
4	1	300	0.899	0.877	0.999	0.946	0.934	1.000	0.987	0.991	1.000
4	1	500	0.863	0.907	0.997	0.927	0.951	1.000	0.971	0.987	1.000
4	1	1000	0.865	0.881	1.000	0.926	0.949	1.000	0.985	0.986	1.000

Second, analogous to Pham et al. (2024), we also see fairly significant distinctions between the simple mean and the weighted aggregate approaches of HMPI. E.g., for the whole sample, Table 5 shows that the productivity decreased significantly in the periods 1990–1995 and 2005–2010, while it increased significantly from 2000 to 2005. However, Table 6 shows a continuous decline in productivity from 2000 to 2019, illustrating that the evolution of the productivity changes over the sample period is different between the unweighted and weighted approaches. Moreover, although both tables show that productivity for the entire sample decreased from 2005 to 2010, Table 5 indicates that the productivity decreased by about 8.90%, while Table 6 indicates it decreased much more, by about 20.18%, when considering

Table 4 continued

p	q	n	0.90			0.95			0.99		
			(i)	(ii)	(iii)	(i)	(ii)	(iii)	(i)	(ii)	(iii)
5	1	20	0.839	0.726	0.827	0.908	0.813	0.891	0.969	0.912	0.956
5	1	50	0.872	0.837	0.951	0.934	0.895	0.980	0.975	0.963	0.994
5	1	100	0.866	0.860	0.988	0.931	0.921	0.998	0.986	0.980	1.000
5	1	200	0.873	0.874	0.999	0.934	0.915	1.000	0.987	0.980	1.000
5	1	300	0.867	0.895	1.000	0.922	0.948	1.000	0.975	0.990	1.000
5	1	500	0.852	0.901	1.000	0.922	0.940	1.000	0.980	0.988	1.000
5	1	1000	0.878	0.908	1.000	0.928	0.953	1.000	0.979	0.993	1.000
7	1	20	0.840	0.742	0.853	0.910	0.813	0.899	0.968	0.927	0.962
7	1	50	0.878	0.836	0.974	0.924	0.899	0.990	0.980	0.967	0.999
7	1	100	0.873	0.865	0.993	0.945	0.924	0.999	0.982	0.981	1.000
7	1	200	0.862	0.883	1.000	0.920	0.930	1.000	0.981	0.984	1.000
7	1	300	0.839	0.896	1.000	0.902	0.939	1.000	0.959	0.979	1.000
7	1	500	0.856	0.897	1.000	0.910	0.935	1.000	0.978	0.984	1.000
7	1	1000	0.862	0.900	1.000	0.923	0.937	1.000	0.973	0.981	1.000

the weight of each individual country. This result suggests the importance of examining both unweighted and weighed HMPI to better understand the change in productivity for a group.

Third, the productivity changes for the developed and developing countries are generally different. E.g., both Tables 5 and 6 indicate that the productivity for the developed countries significantly increased in the periods 1995–2000, 2000–2005 and 2015–2019, and decreased from 2005 to 2010. However, for developing countries, Tables 5 and 6 do not reach a consensus on the significance of the productivity change for all the periods we considered, except for 1995–2000, during which both tables identify a significant decrease in productivity.

Finally, we compare the differences and similarities among the productivity changes measured using HMPI versus those measured using MPI with the results presented in Tables 2 and 3 of Zelenyuk and Zhao (2023). It is worth reminding that the MPI in Zelenyuk and Zhao (2023) is computed relative to the conical hull of the production set (i.e., using the conical technical efficiency) while assuming that the technology may still exhibit VRS. Meanwhile, the HMPI here is measured relative to the production set while assuming that the technology exhibits VRS. To simplify our discussions below, we only focus on the entire sample.

In terms of similarities for the unweighted approach, both Table 5 in our paper and Table 2 in Zelenyuk and Zhao (2023) find that productivity experienced a notable decrease between 1990 and 1995 and an increase from 2000 to 2005. In terms of differences for the unweighted approach, Table 5 in our paper also finds that productivity measured using HMPI significantly decreased from 2005 to 2010 while Table 2 in Zelenyuk and Zhao (2023) finds that productivity measured using MPI significantly increased from 1995 to 2000. In terms of similarities for the weighted approach, both Table 6 in our paper and Table 3 in Zelenyuk and Zhao (2023) find that productivity significantly decreased from 2005 to 2010. In terms of differences for the weighted approach, Table 6 in our paper also finds that productivity measured using HMPI significantly decreased in the periods 2000–2005, 2010–2015, 2015–2019 and 1990–2019, while Table 3 in Zelenyuk and Zhao (2023) finds that productivity change measured using MPI was not significant for all the remaining periods except 2005–2010.

Table 5 Results of the estimation for the simple mean HMPI of countries/regions

Year 1	Year 2	exp ($\widehat{\mathcal{H}}_n$)	Bias-corrected estimate	90% CI		95% CI		99% CI	
Entire sample									
1990	1995	0.8814	0.8815***	0.8295	0.9336	0.8195	0.9435	0.8000	0.9630
1995	2000	1.0296	1.0176	0.9811	1.0541	0.9742	1.0610	0.9605	1.0747
2000	2005	1.0888	1.0712***	1.0315	1.1109	1.0239	1.1185	1.0090	1.1334
2005	2010	0.9449	0.9110***	0.8706	0.9515	0.8628	0.9593	0.8476	0.9744
2010	2015	0.9827	0.9686	0.9345	1.0027	0.9279	1.0092	0.9152	1.0219
2015	2019	1.0105	1.0040	0.9687	1.0393	0.9620	1.0460	0.9488	1.0593
1990	2019	0.9883	0.9526	0.8689	1.0363	0.8529	1.0523	0.8216	1.0836
Developed countries									
1990	1995	1.0256	1.0400	0.9983	1.0817	0.9903	1.0897	0.9747	1.1053
1995	2000	1.1723	1.1675***	1.1389	1.1961	1.1334	1.2016	1.1227	1.2123
2000	2005	1.0675	1.0482***	1.0175	1.0788	1.0117	1.0847	1.0002	1.0962
2005	2010	0.8757	0.7862***	0.7501	0.8223	0.7432	0.8293	0.7297	0.8428
2010	2015	0.9476	0.9122***	0.8777	0.9466	0.8711	0.9533	0.8582	0.9662
2015	2019	1.0415	1.0475***	1.0352	1.0598	1.0329	1.0622	1.0283	1.0668
1990	2019	1.1282	1.0455	0.9682	1.1228	0.9534	1.1376	0.9244	1.1665
Developing countries									
1990	1995	0.8523	0.8651***	0.7941	0.9361	0.7805	0.9497	0.7540	0.9762
1995	2000	0.9742	0.9568*	0.9146	0.9990	0.9066	1.0071	0.8908	1.0229
2000	2005	1.1210	1.1145***	1.0575	1.1715	1.0466	1.1824	1.0252	1.2037
2005	2010	1.0638	1.0832*	1.0255	1.1409	1.0145	1.1519	0.9929	1.1735
2010	2015	1.0369	1.0486	0.9992	1.0980	0.9898	1.1075	0.9713	1.1260
2015	2019	1.0114	1.0057	0.9525	1.0590	0.9422	1.0692	0.9223	1.0891
1990	2019	1.0686	1.0900	0.9554	1.2247	0.9295	1.2505	0.8791	1.3010

Statistical significance (difference from 1) for the bias-corrected estimate of the true simple mean HMPI at the one, five, or ten percent levels is denoted by three, two, or one asterisks, respectively

Table 6 Results of the estimation for the aggregate HMPI of countries/regions

Year 1	Year 2	$\exp(\widehat{\zeta_n})$	Bias-corrected estimate	90% CI		95% CI		99% CI	
Entire sample									
1990	1995	0.9918	0.9780	0.9319	1.0241	0.9230	1.0329	0.9058	1.0501
1995	2000	1.0483	1.0282	0.9695	1.0869	0.9582	1.0981	0.9362	1.1201
2000	2005	1.0234	0.9704*	0.9424	0.9983	0.9370	1.0037	0.9265	1.0142
2005	2010	0.9039	0.7982***	0.7407	0.8558	0.7297	0.8668	0.7081	0.8883
2010	2015	0.9714	0.9182***	0.8738	0.9626	0.8653	0.9711	0.8487	0.9878
2015	2019	0.9927	0.9457*	0.8950	0.9965	0.8853	1.0062	0.8663	1.0252
1990	2019	0.9441	0.7714*	0.5739	0.9690	0.5361	1.0068	0.4622	1.0807
Developed countries									
1990	1995	1.0100	0.9975	0.9431	1.0520	0.9326	1.0625	0.9122	1.0828
1995	2000	1.1295	1.1152***	1.0914	1.1390	1.0868	1.1436	1.0779	1.1525
2000	2005	1.0388	1.0255**	1.0073	1.0437	1.0039	1.0472	0.9971	1.0540
2005	2010	0.9250	0.8959***	0.8299	0.9620	0.8172	0.9746	0.7925	0.9993
2010	2015	1.0073	1.0307	0.9864	1.0749	0.9780	1.0834	0.9614	1.0999
2015	2019	1.0348	1.0442***	1.0362	1.0522	1.0347	1.0537	1.0317	1.0567
1990	2019	1.1559	1.1731*	1.0200	1.3261	0.9907	1.3555	0.9334	1.4128
Developing countries									
1990	1995	0.9758	0.9893	0.8758	1.1029	0.8540	1.1246	0.8115	1.1671
1995	2000	0.9557	0.9166***	0.8782	0.9550	0.8708	0.9624	0.8564	0.9767
2000	2005	1.0891	1.0586	0.9647	1.1525	0.9467	1.1705	0.9115	1.2057
2005	2010	1.0450	1.0114	0.9430	1.0798	0.9298	1.0929	0.9042	1.1186
2010	2015	1.0204	1.0025	0.9664	1.0386	0.9595	1.0455	0.9460	1.0591
2015	2019	1.0076	0.9664*	0.9381	0.9947	0.9327	1.0001	0.9221	1.0107
1990	2019	1.1350	1.0875	0.9135	1.2615	0.8802	1.2948	0.8150	1.3599

Statistical significance (difference from 1) for the bias-corrected estimate of the true aggregate HMPI at the one, five, or ten percent levels is denoted by three, two, or one asterisks, respectively

The differences and similarities between the HMPI and MPI vividly illustrate that examining the productivity change measured using HMPI might provide a different picture from that using MPI, potentially implying different policy implications. Whether it is the HMPI or the MPI that should be used, or perhaps both, is another interesting question (related to economic theory, index numbers theory, context, etc.). What is important, however, is that both are developed on par with each other, in terms of the theoretical framework for estimating and making inferences, and this is what we have aimed to achieve in this paper.

7 Conclusions

In this paper, we address a gap in the existing literature by establishing statistical inference procedures for DEA estimators related to a specific firm's HMPI, the unweighted simple mean HMPI, and the weighted aggregate HMPI. These findings contribute to the ongoing advancement of theoretical results concerning efficiency and productivity estimators derived from non-parametric frontier methods, including Kneip et al. (2015) for the simple mean efficiency, Simar and Zelenyuk (2018) for the aggregate efficiency, Simar and Wilson (2019) for the decompositions of the MPI, Simar and Wilson (2020) for the overall and allocative efficiency, Kneip et al. (2021) for the simple mean MPI, Kneip et al. (2022) for the conical FDH estimators, and Pham et al. (2024) for the aggregate MPI, among others.

We further conduct the MC simulations to assess the effectiveness of the developed CLT for both the simple mean and aggregate HMPI in finite samples. The MC results generally support our developed statistical results. We also utilize a real data set to demonstrate the developed statistical results for the productivity change measured using HMPI by comparing with those measured using MPI.

Future research could focus on deriving asymptotic properties for the decompositions of the HMPI, analogous to Simar and Wilson (2019) for the decompositions of the MPI.[19] Another interesting strand of future work could be to develop an analogous theory for the profitability/allocative Hicks–Moorsteen productivity index proposed by Mayer and Zelenyuk (2019). Another promising direction could be to customize the methods recently introduced by Nguyen et al. (2022), Simar et al. (2023, 2024), and Zelenyuk and Zhao (2023) to possibly enhance the accuracy of finite approximation of the developed CLT for both the simple mean and aggregate HMPI, particularly in scenarios characterized by large dimensions and small sample sizes.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10479-024-06288-8.

Acknowledgements Valentin Zelenyuk acknowledges the support from the Australian Research Council (FT170100401) and The University of Queensland. Shirong Zhao acknowledges the support from National Natural Science Foundation of China (grant number 72401056) and the Liaoning Provincial Department of Education Research Fund (LJKMZ20221602). Feedback from participants of all seminars and conferences where this work was presented (EWEPA, ANZESG, CEPA conference, etc.), as well as from Bao Huang Nguyen, Zhichao Wang and Evelyn Smart is appreciated. We also thank Paul Wilson as the programming codes used in this paper involve some earlier codes from Paul Wilson. These individuals and organizations are not responsible for the views expressed here. These authors contributed equally: Léopold Simar, Valentin Zelenyuk, Shirong Zhao.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions

[19] We note that some recent papers (e.g., Diewert & Fox, 2017) have also provided the decompositions for the HMPI, which might be worth exploring for developing statistical theories.

Declarations

Conflict of interest The authors declare no potential Conflict of interest with respect to the research, authorship, and publication of this article.

References

Abad, A., & Ravelojaona, P. (2017). Exponential environmental productivity index and indicators. *Journal of Productivity Analysis, 48*, 147–166.

Abad, A., & Ravelojaona, P. (2022). A generalization of environmental productivity analysis. *Journal of Productivity Analysis, 57*, 61–78.

Aigner, D., Lovell, C. A. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics, 6*, 21–37.

Amsler, C., Prokhorov, A., & Schmidt, P. (2017). Endogenous environmental variables in stochastic frontier models. *Journal of Econometrics, 199*, 131–140.

Ang, F., Kerstens, K., & Sadeghi, J. (2023). Energy productivity and greenhouse gas emission intensity in Dutch dairy farms: A Hicks–Moorsteen by-production approach under non-convexity and convexity with equivalence results. *Journal of Agricultural Economics, 74*, 492–509.

Aparicio, J., López-Torres, L., & Santín, D. (2018). Economic crisis and public education. A productivity analysis using a Hicks–Moorsteen index. *Economic Modelling, 71*, 34–44.

Arjomandi, A., Valadkhani, A., & O'Brien, M. (2014). Analysing banks' intermediation and operational performance using the Hicks–Moorsteen TFP index: The case of Iran. *Research in International Business and Finance, 30*, 111–125.

Baležentis, T., Blancard, S., Shen, Z., & Štreimikienė, D. (2021). Analysis of environmental total factor productivity evolution in European agricultural sector. *Decision Sciences, 52*, 483–511.

Baležentis, T., Kerstens, K., & Shen, Z. (2022). Economic and environmental decomposition of Luenberger–Hicks–Moorsteen total factor productivity indicator: Empirical analysis of Chinese textile firms with a focus on reporting infeasibilities and questioning convexity. *IEEE Transactions on Engineering Management*.

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

Bjurek, H. (1996). The Malmquist total factor productivity index. *The Scandinavian Journal of Economics, 98*, 303–313.

Briec, W., & Kerstens, K. (2004). A Luenberger–Hicks–Moorsteen productivity indicator: Its relation to the Hicks–Moorsteen productivity index and the Luenberger productivity indicator. *Economic Theory, 23*, 925–939.

Briec, W., & Kerstens, K. (2009). Infeasibility and directional distance functions with application to the determinateness of the Luenberger productivity indicator. *Journal of Optimization Theory and Applications, 141*, 55–73.

Briec, W., & Kerstens, K. (2011). The Hicks–Moorsteen productivity index satisfies the determinateness axiom. *The Manchester School, 79*, 765–775.

Caves, D. W., Christensen, L. R., & Diewert, W. E. (1982). The economic theory of index numbers and the measurement of input, output and productivity. *Econometrica, 50*, 1393–1414.

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

Chen, X., Kerstens, K., & Tsionas, M. (2024). Does productivity change at all in Swedish district courts? Empirical analysis focusing on horizontal mergers. *Socio-Economic Planning Sciences, 91*, 101787.

Chen, X., & Liu, X. (2023). Comparing Malmquist and Hicks–Moorsteen productivity changes in China's high-tech industries: Exploring convexity implications. *Central European Journal of Operations Research, 31*, 1209–1237.

Chen, X., Liu, X., & Zhu, Q. (2022). Comparative analysis of total factor productivity in China's high-tech industries. *Technological Forecasting and Social Change, 175*, 121332.

Daouia, A., & Simar, L. (2007). Nonparametric efficiency analysis: A multivariate conditional quantile approach. *Journal of Econometrics, 140*, 375–400.

Deng, H., Bai, G., Kerstens, K., & Shen, Z. (2023). Comparing green productivity under convex and nonconvex technologies: Which is a robust approach consistent with energy structure? *Managerial and Decision Economics, 44*, 4377–4394.

Diewert, W. E. (1992). Fisher ideal output, input, and productivity indexes revisited. *Journal of Productivity Analysis, 3*, 211–248.

Diewert, W. E., & Fox, K. J. (2017). Decomposing productivity indexes into explanatory factors. *European Journal of Operational Research, 256*, 275–291.

Epure, M., Kerstens, K., & Prior, D. (2011). Technology-based total factor productivity and benchmarking: New proposals and an application. *Omega, 39*, 608–619.

Färe, R., Grosskopf, S., Lindgren, B., & Roos, P. (1994). Productivity developments in Swedish hospitals: A Malmquist output approach. In A. Charnes, W. W. Cooper, A. Y. Lewin, & L. M. Seiford (Eds.), *Data envelopment analysis: Theory, methodology and applications* (pp. 253–272). Dordrecht: Kluwer Academic Publishers.

Färe, R., Grosskopf, S., Norris, M., & Zhang, Z. (1994). Productivity growth, technical progress, and efficiency change in industrialized countries. *American Economic Review, 84*, 66–83.

Färe, R., Grosskopf, S., & Roos, P. (1996). On two definitions of productivity. *Economics Letters, 53*, 269–274.

Färe, R., Mizobuchi, H., & Zelenyuk, V. (2021). Hicks neutrality and homotheticity in technologies with multiple inputs and multiple outputs. *Omega, 101*, 102240.

Farrell, M. J. (1957). The measurement of productive efficiency. *Journal of the Royal Statistical Society A, 120*, 253–281.

Grifell-Tatjé, E., & Lovell, C. K. (1995). A note on the Malmquist productivity index. *Economics letters, 47*, 169–175.

Grifell-Tatjé, E., & Lovell, C. K. (1999). A generalized Malmquist productivity index. *Top, 7*, 81–101.

Grosskopf, S. (1993). Efficiency and productivity. In C. L. H. O. Fried & S. S. Schmidt (Eds.), *The measurement of productive efficiency: Techniques and applications* (pp. 160–194). Oxford: Oxford University Press.

Jin, Q., Kerstens, K., & Van de Woestyne, I. (2020). Metafrontier productivity indices: Questioning the common convexification strategy. *European Journal of Operational Research, 283*, 737–747.

Kerstens, K., Hachem, B. A. M., & Van de Woestyne, I. (2010). Malmquist and Hicks–Moorsteen productivity indices: An empirical comparison focusing on infeasibilities. In: *HUB Research paper 2010/31*. Brussel.

Kerstens, K., Shen, Z., & Van de Woestyne, I. (2018). Comparing Luenberger and Luenberger–Hicks–Moorsteen productivity indicators: How well is total factor productivity approximated? *International Journal of Production Economics, 195*, 311–318.

Kerstens, K., & Van de Woestyne, I. (2014). Comparing Malmquist and Hicks–Moorsteen productivity indices: Exploring the impact of unbalanced vs. balanced panel data. *European Journal of Operational Research, 233*, 749–758.

Kneip, A., Simar, L., & Wilson, P. W. (2015). When bias kills the variance: Central limit theorems for DEA and FDH efficiency scores. *Econometric Theory, 31*, 394–422.

Kneip, A., Simar, L., & Wilson, P. W. (2021). Infernece in dynamic, nonparametric models of production: Central limit theorems for Malmquist indices. *Econometric Theory, 37*, 537–572.

Kneip, A., Simar, L., & Wilson, P. W. (2022). Conical FDH estimators of general technologies, with applications to returns to scale and Malmquist productivity indices. In: *LIDAM discussion paper ISBA 2022/24*. Université catholique de Louvain, Institute of Statistics, Biostatistics and Actuarial Sciences.

Kumbhakar, S. C., Parmeter, C. F., & Zelenyuk, V. (2022). Stochastic frontier analysis: Foundations and advances I. In S. C. Ray, R. Chambers, & S. Kumbhakar (Eds.), *Handbook of production economics* (pp. 331–370). Singapore: Springer.

Kumbhakar, S. C., Parmeter, C. F., & Zelenyuk, V. (2022). Stochastic frontier analysis: Foundations and advances II. In S. C. Ray, R. Chambers, & S. Kumbhakar (Eds.), *Handbook of Production Economics* (pp. 371–408). Singapore: Springer.

Mayer, A., & Zelenyuk, V. (2019). Aggregation of individual efficiency measures and productivity indices. In T. ten Raa & W. H. Greene (Eds.), *The Palgrave handbook of economic performance analysis* (pp. 527–557). Cham: Springer International Publishing.

Mohammadian, I., & Rezaee, M. J. (2020). A new decomposition and interpretation of Hicks–Moorsteen productivity index for analysis of stock exchange companies: Case study on pharmaceutical industry. *Socio-Economic Planning Sciences, 69*, 100674.
Nguyen, H., Simar, L., & Zelenyuk, V. (2022). Data sharpening for improving CLT approximations for DEA-type efficiency estimators. *European Journal of Operational Research, 303*, 1469–1480.
Olesen, O. B., & Ruggiero, J. (2022). The hinging hyperplanes: An alternative nonparametric representation of a production function. *European Journal of Operational Research, 296*, 254–266.
O'Donnell, C. J. (2010). Measuring and decomposing agricultural productivity and profitability change. *Australian Journal of Agricultural and Resource Economics, 54*, 527–560.
Parmeter, C. F., & Zelenyuk, V. (2019). Combining the virtues of stochastic frontier and data envelopment analysis. *Operations Research, 67*, 1628–1658.
Pham, M. D., Simar, L., & Zelenyuk, V. (2024). Statistical inference for aggregation of Malmquist productivity indices. *Operations Research, 72*(4), 1615–1629.
Sala-Garrido, R., Molinos-Senante, M., & Mocholí-Arce, M. (2018). Assessing productivity changes in water companies: a comparison of the Luenberger and Luenberger–Hicks–Moorsteen productivity indicators. *Urban Water Journal, 15*, 626–635.
Schmidt, P., & Sickles, R. C. (1984). Production frontiers and panel data. *Journal of Business & Economic Statistics, 2*, 367–374.
See, K. F., & Li, F. (2015). Total factor productivity analysis of the UK airport industry: A Hicks–Moorsteen index method. *Journal of Air Transport Management, 43*, 1–10.
Seufert, J. H., Arjomandi, A., & Dakpo, K. H. (2017). Evaluating airline operational performance: A Luenberger–Hicks–Moorsteen productivity indicator. *Transportation Research Part E: Logistics and Transportation Review, 104*, 52–68.
Shen, Z., Baležentis, T., & Ferrier, G. D. (2019). Agricultural productivity evolution in China: A generalized decomposition of the Luenberger–Hicks–Moorsteen productivity indicator. *China Economic Review, 57*, 101315.
Shen, Z., Kerstens, K., Baležentis, T. (2023). An environmental Luenberger–Hicks–Moorsteen total factor productivity indicator: Empirical analysis considering undesirable outputs either as inputs or outputs, and attention for infeasibilities. *Annals of Operations Research*, 1–23.
Shen, Z., & Valdmanis, V. (2022). Assessing total factor productivity across Africa: An empirical investigation. *Journal of Productivity Analysis, 58*, 239–253.
Sickles, R. C., & Zelenyuk, V. (2019). *Measurement of productivity and efficiency*. Cambridge: Cambridge University Press.
Simar, L., & Wilson, P. W. (1998). Sensitivity analysis of efficiency scores: How to bootstrap in nonparametric frontier models. *Management Science, 44*, 49–61.
Simar, L., & Wilson, P. W. (2019). Central limit theorems and inference for sources of productivity change measured by nonparametric Malmquist indices. *European Journal of Operational Research, 277*, 756–769.
Simar, L., & Wilson, P. W. (2020). Technical, allocative and overall efficiency: Estimation and inference. *European Journal of Operational Research, 282*, 1164–1176.
Simar, L., & Wilson, P. W. (2023). Another look at productivity growth in industrialized countries. *Journal of Productivity Analysis, 60*, 257–272.
Simar, L., & Zelenyuk, V. (2018). Central limit theorems for aggregate efficiency. *Operations Research, 66*, 137–149.
Simar, L., Zelenyuk, V., & Zhao, S. (2023). Further improvements of finite sample approximation of central limit theorems for envelopment estimators. *Journal of Productivity Analysis, 59*, 189–194.
Simar, L., Zelenyuk, V., & Zhao, S. (2024). Inference for aggregate efficiency: Theory and guidelines for practitioners. *European Journal of Operational Research, 316*, 240–254.
Tsionas, M., Parmeter, C. F., & Zelenyuk, V. (2023). Bayesian artificial neural networks for frontier efficiency analysis. *Journal of Econometrics, 236*, 105491.
Walker, N. L., Styles, D., Gallagher, J., & Williams, A. P. (2021). Aligning efficiency benchmarking with sustainable outcomes in the United Kingdom water sector. *Journal of Environmental Management, 287*, 112317.
Zelenyuk, V. & Zhao, S. (2023). Further improvements of finite sample approximation of central limit theorems for weighted and unweighted Malmquist productivity indices. In: *Working paper series no. WP04/2023*. School of Economics and Centre for Efficiency and Productivity Analysis (CEPA) at The University of Queensland, Australia. https://economics.uq.edu.au/files/42771/WP042023.pdf.