Primal Productivity Indices: Unbalanced vs. Balanced Panel Data

Kristiaan Kerstens

CNRS-LEM, IESEG School of Management (Lille, France).

Ignace Van de Woestyne

Hogeschool Universiteit Brussel (Brussels, Belgium).

NMFS Productivity Workshop, Santa Cruz (CA)
Southwest Fisheries Science Center
11-12 June2012

Outline of the presentation

1. **Introduction**

2. **Definitions of Technology and Primal Productivity Indices**
- **Technology and Distance Functions**
- **Malmquist Productivity Index**
- **Hicks-Moorsteen Productivity Index**
- **Primal Productivity Indices: A Comparison**

3. **Treatments for Unbalanced Panel Data in the Literature**

4. **Data, Methodology, and Empirical Illustration**
- **Data Description**
- **Specifications for the Efficiency Computations**
- **Empirical Results for Productivity Indices**

5. **Conclusions**

1. Introduction
Basic motivation

- Malmquist input and output productivity indices have become popular.

- Hicks-Moorsteen TFP index is becoming a bit popular.

- There is a widespread misconception that these primal productivity indices require balanced panel data and cannot cope with unbalancedness:

1. Hollingsworth & Wildman (2003): "DEA based Malmquist techniques are unable to cope with unbalanced panel estimation procedures".
2. Some popular software to compute these productivity indices cannot handle unbalanced panels: e.g., DEAP or R-package "Nonparaeff" (v. 0.5-3).

- Surprising given that a seminal articles on the Malmquist productivity index points out that an unbalanced panel is possible: "although the index will be undefined for missing observations" (Färe et al. (1994) *AER*: fn 14 on p. 73).

1. Introduction
Main Goal

- Notion of a potential unbalancedness bias due to unplanned missing data is quite standard in the statistical literature (see, e.g., Baltagi & Song (2006) or Frees (2004).
- To the best of our knowledge nobody has so far analysed the extent of the differences between computing primal productivity indices using balanced and unbalanced panel data.
- **Main goal:** This contribution starts to systematically explore the consequences of computing these primal productivity indices using a balanced panel when initially an unbalanced panel data set is available.

2. Definitions of Technology and Primal Productivity Indices
Technology and Distance Functions

- For each time period t, technology is represented by its production possibility set :
  $T^t = \{(x^t, y^t): x^t \text{ can produce } y^t\}$.
- Technology is assumed to satisfy the following conventional assumptions:

(T.1) $(0,0) \in T^t$, $(0, y^t) \in T^t \Rightarrow y^t = 0$.

(T.2) The set $A(x^t) = \{(u^t, y^t) \in T^t : u^t \leq x^t\}$ of dominating observations is bounded $\forall x^t \in \mathbb{R}^n_+$.

(T.3) T^t is closed.

(T.4) $\forall (x^t, y^t) \in T^t$, $(x^t, -y^t) \leq (u^t, -v^t)$ and $(u^t, v^t) \geq 0$ implies that $(u^t, v^t) \in T^t$.

(T.5) T^t is a convex set.

(T.6) $\delta T^t \subseteq T^t, \forall \delta > 0$.

(T.1) Possibility of inaction & no free lunch.
(T.2) Boundedness & (T.3) closedness are mathematical regularity conditions.
(T.4) strong disposal of inputs and outputs.
(T.5) Convexity of technology: linear combinations of activities are feasible.
(T.6) Constant returns to scale.
Notice: (T.5) and (T.6) are not always maintained in this contribution.

2. Definitions of Technology and Primal Productivity Indices
Technology and Distance Functions (2)

- Efficiency is estimated relative to technologies using distance functions or their related efficiency measures.

- Input-oriented Farrell efficiency measure:

$$E_t^i(x^t, y^t) = \inf_{\lambda}\left\{\lambda : (\lambda x^t, y^t) \in T^t, \lambda \geq 0\right\}.$$

- Output-oriented Farrell efficiency measure:

$$E_t^o(x^t, y^t) = \sup_{\theta}\left\{\theta : (x^t, \theta y^t) \in T^t, \theta \geq 1\right\}.$$

- For all $(a,b) \in \{ t, t+1\}^2$, a time-related version of the Farrell input efficiency measure:

$$E_a^i(x^b, y^b) = \inf_{\lambda}\left\{\lambda : (\lambda x^b, y^b) \in T^a\right\}$$

if there is some λ such that $(\lambda x^b, y^b) \in T^a$.

$$E_a^i(x^b, y^b) = +\infty \text{ otherwise.}$$

2. Definitions of Technology and Primal Productivity Indices

Malmquist Productivity Index

- Input-oriented Malmquist productivity index in base period t:

$$M_t^i(x^t, y^t, x^{t+1}, y^{t+1}) = \frac{E_t^i(x^t, y^t)}{E_t^i(x^{t+1}, y^{t+1})}$$

- Input-oriented Malmquist productivity index in base period t+1:

$$M_{t+1}^i(x^t, y^t, x^{t+1}, y^{t+1}) = \frac{E_{t+1}^i(x^t, y^t)}{E_{t+1}^i(x^{t+1}, y^{t+1})}$$

- Geometric mean of a period t and t+1 Input-oriented Malmquist productivity index:

$$M_{t,t+1}^i = \sqrt{M_t^i \cdot M_{t+1}^i}$$

whereby arguments of the functions are suppressed to save space.

- When the geometric mean input-oriented Malmquist productivity index is smaller (larger) than unity, it points to a productivity growth (decline).

2. Definitions of Technology and Primal Productivity Indices
Hicks-Moorsteen Productivity Index

- Hicks-Moorsteen productivity index in base period t:

$$HM_t(x^t, y^t, x^{t+1}, y^{t+1}) = \frac{MO_t(x^t, y^t, y^{t+1})}{MI_t(x^t, x^{t+1}, y^t)}$$

- Hicks-Moorsteen productivity index in base period t+1:

$$HM_{t+1}(x^t, y^t, x^{t+1}, y^{t+1}) = \frac{MO_{t+1}(x^{t+1}, y^{t+1}, y^t)}{MI_{t+1}(x^t, x^{t+1}, y^{t+1})}$$

- Geometric mean of a period t and t+1 Hicks-Moorsteen productivity index:

$$HM_{t,t+1} = \sqrt{HM_t \cdot HM_{t+1}}$$

whereby arguments of the functions are suppressed to save space.

- When the geometric mean Hicks-Moorsteen productivity index is is larger (smaller) than unity, it points to a productivity gain (loss).

2. Definitions of Technology and Primal Productivity Indices

Primal Productivity Indices: A Comparison

Remarks on relative popularity and properties of both primal productivity indices (see also O'Donnell (2008) for details).

- Malmquist productivity index has recently become very popular.
  Hicks-Moorsteen productivity index has found limited use (e.g., O'Donnell (2008) or Zaim (2004)).

- Both ratio-based productivity indices can be related to one another under strict conditions: both coincide under: (i) CRS & (ii) inverse homotheticity.
  Empirical studies comparing both indices are extremely rare: e.g., Bjurek et al. (1998) report minor differences between both indices.
  This limited empirical evidence indicates that the conditions under which both indices coincide do not seem to hold exactly in reality.

- Pitfall of Malmquist: not always a TFP index. Grosskopf (2003) suggests to call it a technology index. In other words, it just measures local technical change, not TFP change.
  Hicks-Moorsteen productivity index has a TFP interpretation (Bjurek (1996) or O'Donnell (2008, 2010)).

2. Definitions of Technology and Primal Productivity Indices

Primal Productivity Indices: A Comparison (2)

- Some of the distance functions constituting the Malmquist productivity index can be undefined when estimated using general technologies.
  Hicks-Moorsteen productivity index satisfies the determinateness axiom.

- Both ratio-based productivity indices can be computed on balanced and unbalanced panel data.
  Distinguish between an infeasibility due to unavailable data and a computational infeasibility.

3. Treatments for Unbalanced Panel Data in the Literature

A List of Proposals

Basic strategy found in literature employing these primal productivity indices consists in making the unbalanced panel somehow balanced.

A variety of strategies can be discerned:

- Simply dropping the observations that are not balanced.

- Sometimes a natural remedy is employed to make the unbalanced panel balanced. *Example:* Backward merger of units: units that merge at some point in time are also treated as merged for the years preceding the merger year.

- More artificial remedies exist to make the initially unbalanced panel balanced. *Example:* Creation of artificial units in an effort to make the panel balanced.

- More elaborate strategies involving some kind of partial balancing: *Example:* Balance on a 2-years by 2-years basis.

- Some proposals to average these productivity indices over a variety of base periods are at least partially motivated by the desire to accommodate unbalanced panel data.

3. Treatments for Unbalanced Panel Data in the Literature

Critiques

General case:

Unbalancedness can occur due to (i) delayed entry, (ii) early exit, or (iii) intermittent nonresponse.

Lack of balance can be:

(i) planned (designed) (for instance, rotating panels), or

(ii) unplanned.

In the unplanned case, non-responses are missing data and a source of bias. This is in particular the case in situations when attrition bias occurs.

Productivity measurement:

- Attrition bias is a known issue, but has not that frequently been reported.
- Unbalancedness is in practice an unknown mix of planned and unplanned. Reason for missing data (i.e., delayed entry, early exit, or intermittent nonresponse) is rarely known.
- If the exact reason is known, then one can measure the contribution of entering and exiting firms to productivity growth (e.g., Griliches & Regev (1995)).

Conclusion: It is useful to at least document the eventual impact of unbalancedness versus balancedness in productivity measurement.

4. Data, Methodology, and Empirical Illustration
A Secondary Data Set

Use a secondary data set in empirical analysis.

Unbalanced panel of 3 years of French fruit producers:
- Based on annual accounting data collected in a survey (Ivaldi et al. (1996)).

- Two outputs: (i) production of apples, and (ii) aggregate of alternative products.

- Prices and quantities of 3 inputs: (i) capital (incl. land), (ii) labour, & (iii) materials.

- 184 farms are available: 130, 135 & 140 have records in 1984, 1985 & 1986 resp.

4. Data, Methodology, and Empirical Illustration
Specifications of Technologies for the Efficiency Computations

Unified algebraic representation of convex and non-convex technologies under CRS or VRS (Briec et al (2004)):

$$T^{\Lambda,\Gamma} = \left\{ (x,y) \in \mathbb{R}_+^{n+p} : y_i \leq \sum_{k=1}^{K} \delta z_k y_{ki},\ (i = 1,\dots,p), \right.$$
$$\left. \sum_{k=1}^{K} \delta z_k x_{kj} \leq x_n,\ (j = 1,\dots,n), z \in \Lambda, \delta \in \Gamma \right\},$$

where $\Lambda \in \{C, NC\}$, with $C = \{z \in \mathbb{R}_+^K : \sum_{k=1}^K z_k = 1\}$ and $NC = \{z \in \mathbb{R}_+^K : \sum_{k=1}^K z_k = 1 \text{ and } \forall k = 1,\dots,K : z_k \in \{0,1\}\}$, and where $\Gamma \in \{CRS, VRS\}$, with $CRS = \mathbb{R}_+$ and $VRS = \{1\}$.

Computing radial input efficiency :

- relative to convex technologies: NLP, or LP.
- relative to non-convex technologies: NLMIP, MIP, LP, or enumeration.

4. Empirical Results

Primal Productivity Indices: Descriptive statistics

		Malmquist				Hicks-Moorsteen			
		Unbalanced		Balanced		Unbalanced		Balanced	
		1984-85	1985-86	1984-85	1985-86	1984-85	1985-86	1984-85	1985-86
$T^{C,CRS}$	n	110	111	92	92	110	111	92	92
	Average	1.1368	1.2213	1.1181	1.2297	1.1793	1.0965	1.1934	1.1070
	Stand. Dev.	0.5439	0.7576	0.5222	0.7978	1.1556	0.7540	1.2234	0.7938
	Min	0.0855	0.1435	0.0854	0.1435	0.3418	0.1913	0.3264	0.1919
	Max	2.9785	5.2625	3.0830	5.2365	11.6472	6.8672	11.6385	6.8672
$T^{C,VRS}$	n	107	108	89	89	110	111	92	92
	Average	1.1369	0.9536	1.1210	0.9432	1.1500	1.1675	1.1611	1.1812
	Stand. Dev.	0.2683	0.3004	0.2811	0.2965	1.1368	0.7237	1.1999	0.7772
	Min	0.5282	0.5583	0.5280	0.5518	0.3318	0.1893	0.3472	0.1925
	Max	1.9068	2.6097	1.9512	2.5140	11.5799	6.4165	11.5746	6.4861
$T^{NC,CRS}$	n	110	111	92	92	110	111	92	92
	Average	1.1429	1.1605	1.1289	1.1707	1.1300	1.1003	1.1359	1.0955
	Stand. Dev.	0.5883	0.5870	0.5711	0.6014	0.8003	0.7421	0.8191	0.7836
	Min	0.1471	0.1252	0.1428	0.1305	0.2507	0.2625	0.2507	0.2829
	Max	4.3777	3.5131	4.3777	3.5803	6.7987	7.0934	7.0040	7.0399
$T^{NC,VRS}$	n	105	107	87	87	110	111	92	92
	Average	1.1116	1.0025	1.1101	1.0161	1.0992	1.1402	1.0995	1.1215
	Stand. Dev.	0.3326	0.3015	0.3579	0.3544	0.6649	0.6579	0.6649	0.6714
	Min	0.4652	0.5062	0.4210	0.4974	0.4015	0.1668	0.4022	0.1668
	Max	2.8566	2.0174	2.8566	2.3650	5.3857	5.7232	5.2018	5.7377

Table 1: Descriptive Statistics for Malmquist and Hicks-Moorsteen Productivity Indices under Various Specifications

4. Empirical Results

Primal Productivity Indices: Descriptive statistics

Conclusions on descriptive statistics:

- Mq and HM disagree on nature of productivity change: Mq points to productivity decline (except under $T^{C,VRS}$), HM always measures productivity growth.
- Descriptive statistics for both indices differ for balanced and unbalanced cases.
- These descriptive statistics seem rather robust across the several specifications of technology (exception under $T^{C,VRS}$).

4. Empirical Results

Primal Productivity Indices: Non-Availabilities & Infeasibilities

		Unbalanced			Balanced		
		1984-85	1985-86	Overall	1984-85	1985-86	Overall
	% na	40.22	39.67	39.95	50.00	50.00	50.00
Malmquist							
$T^{C,CRS}$	% Inf	0.00	0.00	0.00	0.00	0.00	0.00
$T^{C,VRS}$	% Inf	1.63	1.63	1.63	1.63	1.63	1.63
$T^{NC,CRS}$	% Inf	0.00	0.00	0.00	0.00	0.00	0.00
$T^{NC,VRS}$	% Inf	2.72	2.17	2.45	2.72	2.72	2.72
Hicks-Moorsteen							
$T^{C,CRS}$	% Inf	0.00	0.00	0.00	0.00	0.00	0.00
$T^{C,VRS}$	% Inf	0.00	0.00	0.00	0.00	0.00	0.00
$T^{NC,CRS}$	% Inf	0.00	0.00	0.00	0.00	0.00	0.00
$T^{NC,VRS}$	% Inf	0.00	0.00	0.00	0.00	0.00	0.00

Table 2: Malmquist and Hicks-Moorsteen Productivity Indices under Various Specifications: Non-Availabilities ("na") and Infeasibilities ("Inf")

4. Empirical Results

Primal Productivity Indices: Non-Availabilities & Infeasibilities

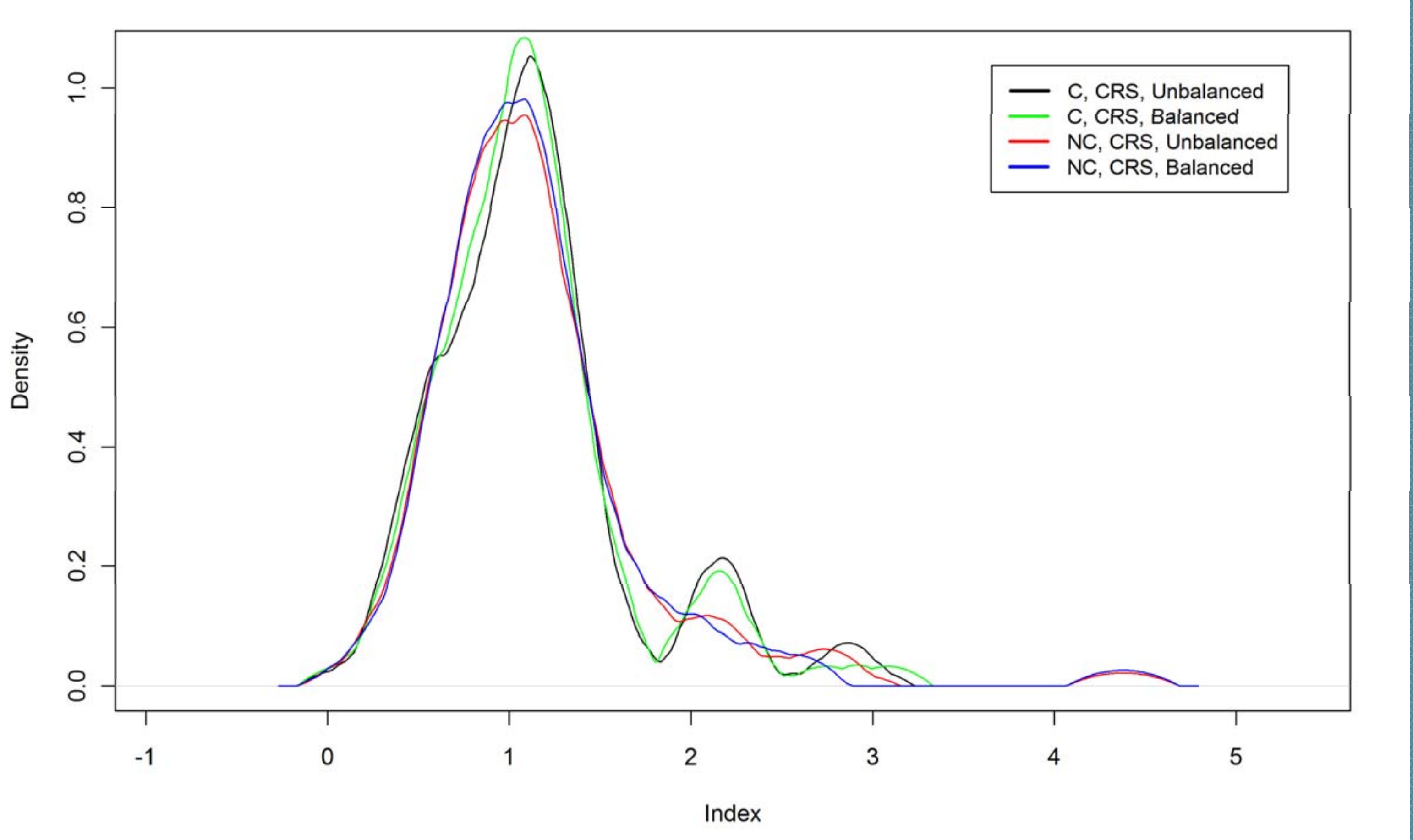
Conclusions on non-availabilities & infeasibilities :

- Infeasibilities due to unavailable data: 50% in balanced case, while around 40% depending on the exact year in unbalanced case. Gain of about 10% in information included in the estimates.
- % computational infeasibilities in Mq is rather stable when comparing balanced and unbalanced cases.
- HM index has no computational infeasibilities.

5. Empirical Results

Balanced and unbalanced Malmquist index for 1984-85: Densities

Figure 1: Kernel Density Estimates of Balanced and Unbalanced Malmquist Index (1984-85) under $T^{C,CRS}$ and $T^{NC,CRS}$



5. Empirical Results

Primal Productivity Indices: Li-test results

		Malmquist		Hicks-Moorsteen	
		1984-85	1985-86	1984-85	1985-86
$T^{C,CRS}$	z-value	-1.1264	-0.9661	-1.0471	-0.9330
	p-value	0.1300	0.1670	0.1475	0.1754
$T^{C,VRS}$	z-value	-1.0497	-0.8951	-1.0001	-0.9975
	p-value	0.1469	0.1854	0.1586	0.1593
$T^{NC,CRS}$	z-value	-1.0439	-1.0173	-0.9673	-0.8615
	p-value	0.1483	0.1545	0.1667	0.1945
$T^{NC,VRS}$	z-value	-0.8854	-0.7358	-0.9764	-0.8456
	p-value	0.1880	0.2309	0.1644	0.1989

Table 3: Li-test Results of Density Comparison between Balanced and Unbalanced Malmquist and Hicks-Moorsteen Productivity Indices under Various Specifications

Conclusion from Li-test comparing balanced and unbalanced results:
Null hypothesis of equality of both balanced and unbalanced distributions cannot be rejected.

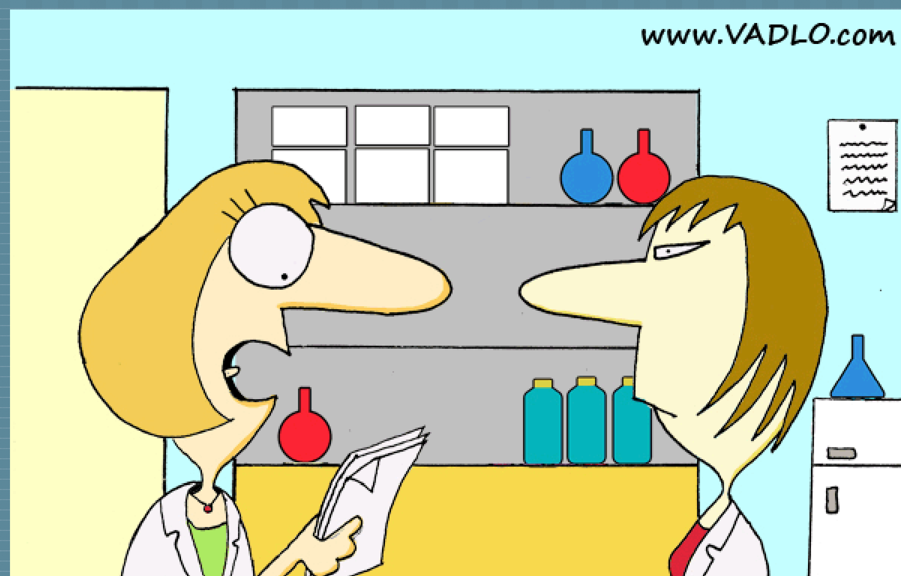
5. Conclusions

- What has been achieved?
  - This contribution is -to the best of our knowledge- the first to empirically illustrate the differences in between using either unbalanced or balanced panel data when computing frontier estimates for the primal Malmquist and Hicks-Moorsteen productivity indices.
  - Data of French fruit producers yields differences between balanced and unbalanced productivity indices, but these turn out not to be significant.
- General perspectives:
  - Be cautious with balancing unbalanced panel data.
  - Need for deeper study on attrition bias and productivity (especially using these widely used primal productivity indices).

The End

Thanks for your attention

Any questions???



"Editor says the manuscript would serve some purpose if it were written on toilet paper."