Accounting and Business Research
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rabr20

Statistical inference using the T index to quantify the level of comparability between accounts

Professor Ross H. Taplin [a]

[a] School of Accounting, Curtin Business School, Curtin University, GPO Box U1987, Perth, Australia, 6845 E-mail:
Published online: 04 Jan 2011.

Statistical inference using the T index to quantify the level of comparability between accounts

Professor Ross H. Taplin*

Abstract — The extent to which the accounts of companies are comparable is considered important to users and regulators. However, prior research has been restricted by a lack of appropriate statistical methods for testing comparability indices. This has made it difficult to assess the true level of comparability from sample data and to test research hypotheses such as whether the level of comparability (a) differs by policy, (b) differs by country, and (c) changes over time.

This paper fills this gap by exploring the statistical properties of the *T* index. The *T* index generalises the *H*, *C*, *I* and various modifications of these indices and represents a unified framework for the measurement of the extent to which the accounts of companies are comparable. Formulae for the bias and standard error for any index under this framework are provided and proved. The bias is shown to equal zero or be negligible in most practical situations. Using historical data, the standard error is used to illustrate the accuracy with which comparability is estimated and to perform formal statistical inference using confidence intervals and p-values. Furthermore, the sampling distribution of the *T* index is assessed for normality. Implications for research design and sample size determination are also discussed.

Keywords: Herfindahl H index; C index; harmony; standardisation

1 Introduction

The *T* index was introduced by Taplin (2004) to quantify the degree to which the accounts of companies are comparable. It is easily interpreted as the probability that two randomly selected companies have accounts that are comparable, or as the average comparability of pairs of companies. The *T* index is a generalisation of the *H*, *I* and *C* indices introduced by van der Tas (1988), and is a framework containing countless individual indices. Many authors have made minor modifications to the basic *H*, *I* and *C* indices to deal with issues such as non-disclosure of the accounting method by a company and many of these are also special cases of the unified approach described by the *T* index. For details of the history of these indices, references to these modifications, literature using these indices, and literature that considers alternative definitions of harmony or related ideas of harmonisation, uniformity and standardisation, the reader is referred to Taplin (2004), the literature review Ali (2005), or Cole et al. (2008), as well as the references contained within these articles.

This paper uses the term 'comparability' in place of the more traditional term 'harmony' used in Taplin (2004) and by papers going back to van der Tas (1988). This is to avoid confusion over terms harmonisation, standardisation and uniformity that potentially have different meanings and positive or negative connotations to different readers (Tay and Parker, 1990). Cole et al. (2008) summarise the changing landscape concerning different perspectives on these terms and on the uniformity-flexibility dilemma when it comes to the extent to which all companies should be forced to use the same method on one extreme, or allowed to use any method they choose on the other extreme. Barth et al. (1999) use a mathematical model to investigate, under several assumptions, the impact of changes such as harmonising domestic regulations in two countries on characteristics such as security market performance. They conclude from their theoretical model that harmonisation is not necessarily desirable.

This paper is concerned with the measurement of the extent to which the actual accounts of companies are comparable. This is important regardless of philosophical perspectives or opinions concerning the uniformity-flexibility continuum and regardless of current regulations because there will always be an interest in knowing the extent to which the accounts prepared by companies are comparable. Comparability is important in concepts such as

*The author is at the School of Accounting, Curtin Business School at Curtin University in Perth, Australia. He acknowledges the constructive comments of the reviewers and the editor that have led to significant improvements to the presentation of this paper.

Correspondence should be addressed to: Professor Ross H. Taplin, School of Accounting, Curtin Business School, Curtin University, GPO Box U1987, Perth 6845, Australia. E-mail: R.Taplin@curtin.edu.au.

Note: The Harmoniser software that calculates the T index is available free of charge from the author.

This paper was accepted for publication in July 2009.

harmonisation, standardisation and uniformity but this paper makes no statement about the preferred position on the uniformity-flexibility continuum. This paper specifically concerns statistical sampling issues when measuring the extent to which company accounts are comparable.

The *T* index is flexible concerning what it means for the accounts of two companies to be comparable. For example, two companies both using FIFO would normally be considered comparable. Companies not disclosing their method, using a combination of methods (FIFO for some inventory and average cost for other inventory) or using multiple methods (results using FIFO for all inventory in addition to results of using average cost for all inventory), for example, must have their comparability defined in a sensible way. Consider a company using FIFO for some of its inventory and average cost for its other inventory and a second company using FIFO for all its inventory. Simple indices prior to the *T* index would consider the accounts of these companies to be completely non-comparable because they are defined to be using different accounting methods. With the *T* index, these accounts can be defined to be completely comparably or partially comparable. For example, if two-thirds of the inventory of the first company was costed using FIFO these two companies might be defined as two-thirds comparable (see Astami, 2006) for further application of partial comparability). Alternatively, the accounts of these companies may be considered completely comparable with the *T* index if they each use the most appropriate method for their circumstances and their type of inventory.

Similarly, consider two companies using straight line for depreciation but one company depreciates over three years while the other depreciates over five years for the same type of asset. Simple indices prior to the *T* index were forced to consider these as the same method, and therefore completely comparable with each other, or different methods, and therefore completely non-comparable with each other. The *T* index, however, allows the level of comparability to be partial (a value between zero representing completely non-comparable and one representing completely comparable). Furthermore, if straight line was used for an asset that depreciated non-linearly there is a strong case that comparability is weak with the accounts of a company correctly depreciating along a straight line. In this instance it may be necessary to define two different accounting methods, both methods are for companies using straight line depreciation however one is for companies where this is appropriate and the other is for companies where it is inappropriate. This flexibility makes the *T* index applicable in many situations but also demands careful reasoning and justification for an appropriate definition of comparability rather than just using a simple but convenient index.

While not the topic of this paper, the *T* index is sufficiently flexible to allow very different concepts of comparability. For example, two companies that both use straight line depreciation over the same time period for the same asset, when the asset actually depreciates exponentially, may be considered non-comparable because both are unreliable or inaccurate assessments of the companies. However, if we separate the desirable qualitative characteristics of reliability from comparability for company accounts, we could define these two companies to be completely comparable (but both unreliable). For example, if both companies have identical accounts but both over-estimate their depreciation (by the same amount) we correctly conclude these companies have identical accounts so our comparability is not compromised even though the reliability is low for each company. Reliability is not considered in this paper although it is possible to require reliability before accounts are defined to be comparable. The focus of this paper centres on techniques of statistical inference for comparability indices for any definition of comparability because the results in this paper hold for any comparability index within the *T* index framework. Furthermore, as revealed by correspondence with reviewers, the concept of a sensible definition of comparability involves subjective opinions and can change over time and with the circumstances in which it is applied.

The adoption of International Financial Reporting Standards (IFRS) is expected to enhance comparability of accounts but many countries have not yet agreed to follow IFRS and within IFRS policy choice is still allowed. Nobes (2006) argues international differences will persist under IFRS and proposes a research agenda with many research hypotheses concerned with the extent to which company accounts in different countries are comparable. Cole et al. (2008) argue that differences in the application of IFRS will lead to persistent lack of comparability, and in their review concluded the *T* index was the most appropriate methodology for measuring comparability. Nevertheless, without appropriate statistical inference techniques to develop and test research hypotheses (or just quantify the accuracy of comparability estimated from samples), research using the *T* index, such as Astami et al. (2006) and Cole et al. (2008), is

hampered. More recently, Cairns et al. (2009) successfully applied the results outlined in this paper to investigate changes in comparability for UK and Australian companies around the time of adoption of IFRS.

The need for statistical inference in research is well understood. Knowing a sample estimate of a population quantity is arguably of no value if a measure of the accuracy of this estimate cannot also be provided. This was recognised very early in the development of indices for accounting comparability, as expressed by Tay and Parker (1990) '[t]he main problem with concentration indices is that no significance tests have been devised to indicate how trivial or significant (statistically) variations in index values are'. Taplin (2003) responded by providing formulae for the bias and standard error of the H index and C index. Unfortunately, these formulae only apply to two specific indices that are suitable for specific research questions and only for data with specific characteristics. For example, the formulae in Taplin (2003) do not apply if the accounts of a company are comparable with companies using several different accounting methods or if comparisons between companies in different countries are required. The T index was developed precisely to provide a framework whereby indices with desirable characteristics could be chosen from within a unified framework.

This paper therefore adds to Taplin (2004) by providing the necessary details to enable statistical inference to be performed with any index within the T index framework. This will greatly enhance research using the T index to quantify the level of comparability.[1]

The rest of this paper is structured as follows. Section 2 contains an overview of the T index. Section 3 provides formulae for the bias and standard error of the T index and for the special cases known as the H, I, C and between country C and within country C indices. This forms the major mathematical results of the paper. Herrmann and Thomas (1995) reported an example using data on nine measurement practices from eight countries but without any analysis of statistical significance. In Sections 4 to 7 we provide full statistical results using this data: Section 4 overall indices; Sections 5 and 6 comparisons of fairness and legalistic countries with different treatments of non-disclosure, and Section 7 the two-country I index. Section 8 investigates empirically the sampling distribution of the T index. Section 9 shows how the formula for the standard error for the T index can be used to perform sample size calculations. Section 10 contains some concluding discussion. The Appendix contains the formulae to calculate the standard error of the T index and a mathematical derivation of the formulae for the bias and standard error.

2. The T index

The T index is easily interpreted as the probability that two randomly selected companies have accounts that are comparable, or as the average comparability of pairs of companies. This requires defining the comparability between pairs of accounting methods and how the random sampling of companies is performed. This is achieved by specifying coefficients α_{kl} and β_{ij} respectively, with different choices of these coefficients resulting in different specific indices from within the T index framework. The α_{kl} specify the level of comparability between accounting methods k and l. For example, whether companies using FIFO for inventory are comparable to companies not disclosing their method. The β_{ij} specify the way companies are randomly selected. For example, requiring the two selected companies for comparison to be from different countries results in a measure of international comparability.

The general formula for T is given by

$$T = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{M} \sum_{l=1}^{M} \alpha_{kl} \beta_{ij} p_{ki} p_{lj} \tag{1}$$

where

α_{kl} is the coefficient of comparability between accounting methods k and l,

β_{ij} is the weighting for the comparison between companies in countries i and j,

p_{ki} is the proportion of companies in country i that use accounting method k,

p_{lj} is the proportion of companies in country j that use accounting method l,

and there are N countries (labelled 1 to N) and M accounting methods (labelled 1 to M).

As discussed in the introduction, an accounting method is not necessarily equivalent to a procedure such as straight line depreciation because we may

[1] Determinant studies, such as Jaffar and McLeay (2007), provide a different but complementary approach to indices considered in this paper. Indices of comparability are valuable because they quantify comparability directly, emphasise whether companies are comparable, often concentrate on country differences which are commonly found to be the major determinant of policy choice and can be used to investigate which countries, regions or industries contain companies whose accounts are highly comparable. Finally, an index of harmony is a concise summary statistic that is a useful addition to research findings in a similar way to a correlation coefficient or regression R-squared value is, even when these summary statistics are not the major focus of the research.

Figure 1
Options for the T index under four criteria
Company/country weightings
(1a) companies are weighted equally, $b_i = n_i/n$, where n_i is the number of companies from country i in the sample and n is the total number of companies in the sample, so b_i is the proportion of companies in the sample from country i. This means a country receives weight proportional to the number of companies sampled from that country.
(1b) countries are weighted equally, $b_i = 1/N$, where N is the number of countries,
(1c) countries are weighted according to the total population number of companies in each country, $b_i = u_i / \sum_{i=1}^{N} u_i$ where u_i is the total number of companies in country i (for example, the total number of companies listed on the stock exchange rather than the number of companies in the sample).

International focus
(2a) overall, $\beta_{ij} = b_i b_j$,
(2b) within country, $\beta_{ij} = 0$ if $i \neq j$ and when $i = j$, $\beta_{ii} = b_i^2 / \sum_{i=1}^{N} b_i^2$,
(2c) between country, $\beta_{ij} = b_i b_j / \sum_{i=1}^{N} \sum_{j \neq i} b_i b_j$ if $i \neq j$ and when $i = j$, $\beta_{ii} = 0$, where the summation for j is over all countries 1 to N except for country i.

Multiple accounting policies
(3a) multiple accounting policies are not allowed, $\alpha_{kl} = 0$ if $k \neq l$,
(3b) multiple accounting policies are allowed if completely comparable, $\alpha_{kl} = 1$ when methods k and l are completely comparable and $\alpha_{kl} = 0$ when they are completely incomparable,
(3c) multiple accounting policies are allowable with fractional comparability, α_{kl} takes a value on the continuum from 0 (completely incomparable) to 1 (completely comparable).

Non-disclosure
Here it is assumed non-disclosure is the last accounting method M.
(4a) not applicable, companies who do not disclose a method are removed from the sample,
(4b) comparable to everything, $\alpha_{kM} = \alpha_{Ml} = \alpha_{MM} = 1$ for all accounting methods k and l,
(4c) comparable to nothing, $\alpha_{kM} = \alpha_{Ml} = \alpha_{MM} = 0$ for all accounting methods k and l,
(4d) comparable to the standard (or default) method s, $\alpha_{ks} = \alpha_{kM}$, $\alpha_{sl} = \alpha_{Ml}$ for all k and l.

require the same period of depreciation, type of asset or suitability of this method in the circumstances before we define companies to be using the same method. Thus the term accounting method is used as a generic label as it has been in the past literature on comparability indices. In particular, non-disclosure of a method or non-applicability of any method are defined to be accounting methods.

In order to ensure that T is between 0 (no two companies are comparable) and 1 (all companies are comparable with each other) we require the α_{kl} and β_{ij} to be between 0 and 1 (inclusive) and that the β_{ij} sum to 1. The α_{kl} define the comparability between accounting methods k and l, with $\alpha_{kl} = 0$ specifying that the two accounting methods are completely non-comparable and $\alpha_{kl} = 1$ specifying that the two accounting methods are completely comparable. The β_{ij} specify weights for comparisons between companies from countries i and j. For example, β_{ii} specifies the weight given to the comparability of companies from country i while β_{ij} ($i \neq j$) specifies the weight given to comparisons of companies from country i with companies from country j.

Although the coefficients α_{kl} and β_{ij} can be selected very generally to suit the particular data and research questions under analysis, in practice they can be determined by selecting from some intuitive options under four criteria. The four criteria and their respective options, summarised in Figure 1, are discussed in detail in Taplin (2004) and Taplin (2006). The first two criteria define the β_{ij} and the last two criteria define the α_{kl}.

Thus the T index represents an extremely flexible framework containing an uncountable number of specific indices, including many simpler indices. The H index equals the T index under options 1a2a3a4a but is usually applied to a single country.

Table 1
Data X_{ki} format for an example taken from Herrmann and Thomas (1995) for inventory costing

Accounting method (k)	*Country (i)*								*Total*
	Den (i =1)	*Ire* (i =2)	*Neth* (i =3)	*UK* (i =4)	*Bel* (i =5)	*Fra* (i =6)	*Ger* (i =7)	*Por* (i =8)	
FIFO (k =1)	9	8	9	7	2	5	0	2	42
LIFO (k =2)	0	0	1	0	6	2	15	0	24
Average (k =3)	2	0	0	2	8	11	3	9	35
Combination (k =4)	1	0	1	3	3	6	1	8	23
Not disclosed (k =5)	18	16	19	18	4	6	11	1	93
Sample size (n_i)	30	24	30	30	23	30	30	20	217

The H and T indices employ sampling with replacement when selecting two companies for comparison while the C index employs sampling without replacement. Taplin (2004) provides reasons why sampling with replacement is preferable, but in practice both index values will be almost identical unless sample sizes are very small. Similarly, the within country C index gives almost identical values to the T index under options 1a2b3a4a. The between-country C index and T index under options 1a2c3a4a give identical results since sampling with or without replacement are equivalent when only one company is selected from each country. For two countries, the I index equals the between-country C index (Morris & Parker, 1998) and T index (options 1a2c3a4a) but for more than two countries the I index is not a special case of the T index. See Taplin (2004) for a review of undesirable properties of the I index with more than two countries.

This paper takes illustrative data and examples from Herrmann and Thomas (1995) to illustrate the methods in this paper. Their study is well known as a comparative evaluation, has moderate sample sizes, the data is already published so results in this paper can be verified, their examples include issues such as non-disclosure and combination methods, and the effect of different methods on conclusions can be seen more clearly because the same data has been examined in the literature using different methods.

Table 1 illustrates the format of the data required for the T index using an example for inventory costing from Herrmann and Thomas (1995). In this example there are $N = 8$ countries and $M = 5$ accounting methods for inventory costing (non-disclosure of the treatment of inventory costing is the fifth 'method'). Countries are referred to by the index i (i=1 to N) and accounting methods are referred to by the index k (k=1 to M). The data consists of the number of companies in country i using method k, denoted X_{ki} and displayed in Table 1. Sample sizes for each country, denoted n_i are also provided.

We illustrate the calculation of the T index under options 1a2a3a4a. This means that companies are weighted equally, all companies regardless of country are compared (overall international focus), multiple accounting policies do not exist and companies not disclosing a method are removed (leaving only $M = 4$ methods). We use these options for illustrative purposes only and do not suggest they are the most appropriate options for this data. In this case Equation (1) is a summation of $4 \times 4 \times 8 \times 8 = 1,024$ terms, although at least 75% (or $12 \times 8 \times 8 = 768$) of these terms equal zero because 12 of the 16 α_{kl} equal zero. Nevertheless, the T index generally contains a large number of terms and a systematic approach is required. This is provided by the observation that Equation (1) can be written as

$$T = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{M} \sum_{l=1}^{M} \alpha_{kl} \beta_{ij} p_{ki} p_{lj} = \sum_{i=1}^{N} \sum_{j=1}^{N} \beta_{ij} T_{ij}$$

where

$$T_{ij} = \sum_{k=1}^{M} \sum_{l=1}^{M} \alpha_{kl} p_{ki} p_{lj}$$

is the two-country index quantifying the level of harmony between country i and country j and β_{ij} are the weights assigned to the T_{ij} when computing the weighted average. With options 3a and 4a the α_{kl} equal zero when $k \neq l$ and the α_{kk} equal one, so T_{ii} simplifies to the H index for country i and T_{ij} ($i \neq j$) simplifies to the I index for countries i and j. These values are provided in Table 2 with the calculation of T_{11} and T_{31} illustrated beneath the table.

The value of $T = \sum_{i=1}^{N} \sum_{j=1}^{N} \beta_{ij} T_{ij}$ is a weighted average of the T_{ij} values in Table 2. The weights in this average under options 1a and 2a are $\beta_{ij} = b_i b_j = n_i n_j / n^2$ where n_i is the sample size

Table 2
T_{ij} under options 3a4a (with standard errors in parentheses). Diagonal entries T_{ii} equal H indices for country i and off-diagonal entries T_{ij} ($i \neq j$) equal two-country I (or equivalently between-country C) indices. The T index equals a weighted sum of the T_{ij} values.

	Den	*Ire*	*Neth*	*UK*	*Bel*	*Fra*	*Ger*	*Por*
Denmark	0.60	0.75	0.62	0.49	0.16	0.25	0.03	0.19
(n_1 = 12)	(0.14)	(0.13)	(0.13)	(0.11)	(0.06)	(0.06)	(0.02)	(0.07)
Ireland	0.75	1.00	0.82	0.58	0.11	0.21	0.00	0.11
(n_2 = 8)	(0.13)	(0.00)	(0.12)	(0.14)	(0.07)	(0.08)	(0.00)	(0.07)
Netherlands	0.62	0.82	0.69	0.50	0.13	0.20	0.08	0.12
(n_3 = 11)	(0.13)	(0.12)	(0.16)	(0.12)	(0.06)	(0.07)	(0.07)	(0.06)
UK	0.49	0.58	0.50	0.43	0.17	0.26	0.04	0.25
(n_4 = 12)	(0.11)	(0.14)	(0.12)	(0.11)	(0.05)	(0.05)	(0.03)	(0.06)
Belgium	0.16	0.11	0.13	0.17	0.31	0.28	0.32	0.28
(n_5 = 19)	(0.06)	(0.07)	(0.06)	(0.05)	(0.06)	(0.05)	(0.08)	(0.06)
France	0.25	0.21	0.20	0.26	0.28	0.32	0.15	0.34
(n_6 = 24)	(0.06)	(0.08)	(0.07)	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)
Germany	0.03	0.00	0.08	0.04	0.32	0.15	0.65	0.10
(n_7 = 19)	(0.02)	(0.00)	(0.07)	(0.03)	(0.08)	(0.05)	(0.12)	(0.05)
Portugal	0.19	0.11	0.12	0.25	0.28	0.34	0.10	0.41
(n_8 = 19)	(0.07)	(0.07)	(0.06)	(0.06)	(0.06)	(0.05)	(0.05)	(0.06)

Examples:
From the number of companies using each method in each country (see Table 1):
$T_{11} = (9/12)^2 + (0/12)^2 + (2/12)^2 + (1/12)^2 = 0.60$
$T_{13} = (9/12)(9/11) + (0/12)(1/11) + (2/12)(0/11) + (1/12)(1/11) = 0.62$
The T index is given by the weighted average of $T = \sum_{i=1}^{N} \sum_{j=1}^{N} \beta_{ij} T_{ij}$ where β_{ij} are defined by options 1 and 2. Under options 1a2a (companies weighted equally and overall international focus), $\beta_{ij} = n_i n_j / n^2$ for all values of i and j so T equals the sum of 64 terms as follows (first three terms shown only):
$T = (12 \times 12/124^2) \times 0.60 + (8 \times 12/124^2) \times 0.75 + (12 \times 11/124^2) \times 0.62 + ... = 0.27$

for country i (see column 1 of Table 2) and $n = 124$ is the total sample size. The first three terms of this sum are provided beneath Table 2. The resulting value of $T = 0.27$ (options 1a2a3a4a) results from the large spread of companies using different methods and implies there is only a 27% chance of two randomly selected companies having accounts that are comparable. Under options 1b2a3a4a the weights β_{ij} all equal 1/64 and $T = 0.31$ is a simple average of the 64 T_{ij} in Table 2.

If we consider the accounts of non-disclosing companies to be not comparable with accounts of all other companies (option 4c instead of 4a) then $T = 0.087$. In this case $M = 5$ since non-disclosure is included as an accounting method and the high level of non-disclosure results in a lower value of the T index. Note the T_{ij} will be lower than those in Table 2 due to the non-comparability of non-disclosure and the β_{ij} will differ since the sample sizes n_i (and hence n) will be higher.

Many international accounting studies prefer to examine the comparability of companies between different countries (option 2c), a property of the I index and between-country C index. If, as with the I index, we give each country equal weight (option 1b), β_{ij} equals 1/56 when $i \neq j$ and $\beta_{ii} = 0$. Then $T = 0.27$ (option 1b2c3a4a) is a simple average of the off-diagonal entries in Table 2. If companies are weighted equally (option 1a2c3a4a) we obtain the between-country C index value of $T = 0.24$. If non-disclosing companies are considered non-comparable to all other companies (option 4c) these values for the T index are 0.079 and 0.076 respectively.

The choice of α_{kl} and β_{ij} in any application requires careful consideration and justification. These examples are for illustration only and we do not claim any of the above choices are optimal. Indeed, calculating and reporting values of the T index under different assumptions or options, as above, is recommended. The flexibility of the T index provides a unified framework for comparison of indices, encourages careful thought of the appropriate index for a particular problem, enhances investigations into the sensitivity of conclusions to the choice of index, and opens up

possibilities of new indices tailor made for a specific problem.

3. Statistical inference for the T index

The T index given by Equation (1) is typically calculated using sample data consisting of proportions p_{ki} equal to the proportion of companies in the sample of companies from country i that use accounting method k. When considering statistical inference for the T index it is important to distinguish between this index based on a sample of companies and the corresponding index based on the population of all companies from these countries. We refer to the latter as the population T index, denoted T_p. It is given by

$$T_p = \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{M} \sum_{l=1}^{M} \alpha_{kl}\beta_{ij}\pi_{ki}\pi_{lj} \tag{2}$$

where

π_{ki} equals the proportion of companies using accounting method k out of all the companies in the population of companies from country i, and

π_{lj} equals the corresponding proportion for method l and country j.

In practice, it is not possible to include all companies in all countries in our sample and hence the sample T index is only an estimate of the population index T_p. Here we consider the statistical properties of the sample T index and in particular how accurate it is as an estimate of the corresponding index in the population.

3.1. The bias of the T index

The bias of an estimate is defined as the difference between the expected value of the sample estimate and the quantity being estimated, and hence equals zero only if the mean of the sampling distribution equals the quantity being estimated.

The bias of the T index is derived in the Appendix to equal the summation in Equation (3) (boxed below).

$$\text{bias}(T) = \sum_{i=1}^{N} \sum_{k=1}^{M} \alpha_{kk}\beta_{ii}\pi_{ki}/n_i - \sum_{i=1}^{N} \sum_{k=1}^{M} \sum_{l=1}^{M} \alpha_{kl}\beta_{ii}\pi_{ki}\pi_{li}/n_i. \tag{3}$$

Although this bias appears a complicated expression that can be positive, negative or zero, there are a few important characteristics that we now discuss.

The bias of T will equal zero if the β_{ii} all equal zero, and in particular for any T index under option (2c) utilising a between-country focus. This is an important special case because international studies often focus on comparisons between countries. With this international focus, we now know that the value of the T index calculated from a random sample will, on average, equal the value of the T index calculated from the entire population. In particular, this result proves the between country C index of Archer et al. (1995) and the two-country I index of van der Tas (1988) are both unbiased.

Further insights into the bias of the T index can be obtained by writing Equation (3) as

$$\text{bias}(T) = \sum_{i=1}^{N} \frac{\beta_{ii}}{n_i}(D_i - T_i). \tag{4}$$

where

$$D_i = \sum_{k=1}^{M} \alpha_{kk}\pi_{ki}$$

and

$$T_i = \sum_{k=1}^{M} \sum_{l=1}^{M} \alpha_{kl}\pi_{ki}\pi_{li}.$$

In many applications of the T index an accounting method will be considered completely comparable with itself (so the α_{kk} will all equal 1) except where one accounting method represents non-disclosure in which case this α_{kk} may be zero. In this case, D_i equals the proportion of companies in the population of country i that disclose their accounting method. When all companies disclose their accounting method, D_i will typically equal 1. Although uncommon in the past literature (Astami et al. (2006) is an exception) α_{kk} can be between 0 and 1 due to partial disclosure, as discussed in the Introduction.

The T_i measure national comparability, the level of comparability for companies from country i, and are based on the population of all companies from country i. For example, if α_{kl} equals 1 when $k = l$ and equals 0 when $k \neq l$, then T_i equals the value of the Herfindahl H index applied to all companies from country i.

From Equation (4) several observations can be made concerning the bias of the T index. First, the bias is rarely negative because it is unlikely T_i will be greater than D_i for any country. If $\alpha_{kl} \leq \alpha_{kk}$ for all values of k and l, which is plausible in practice because it states that an accounting method k is at

least as comparable with itself than with any other accounting method, then $T_i \leq D_i$ and so from Equation (4) the bias is not negative.[2]

Second, since D_i and T_i are both between 0 and 1, their difference must be at most 1 in magnitude. Hence the magnitude of the bias can not be greater than $\sum_{i=1}^{N} \beta_{ii}/n_i$. In practice this is a useful upper bound because it can be calculated prior to data collection since it only depends on the sample size and the international focus to be used for the T index. This implies that the bias will be negligible in large sample sizes.

In summary, the bias will be zero for a between country focus and small if the within country weighting given by $\sum_{i=1}^{N} \beta_{ii}$ is small or if the sample sizes for the countries are all large. For most plausible indices and practical data, the bias will be zero or negligible.

3.2. The standard error for the T index

Formulae to calculate the standard error of the T index are provided in the Appendix. Since the formulae are complicated and not particularly intuitive they are presented in a format suitable for implementation rather than to provide intuitive insights. Examples in the following sections will provide insights into the magnitude of the standard error in different situations.

Instead, Table 3 provides formulae for the variance of the special cases of the T index corresponding to simple H, I, and between-country and within-country C indices. The overall C index gives values slightly different to the H index and T index (option 1a2a3a4a) due to differences between sampling with or without replacement, but these differences are negligible unless sample sizes are very small. Similarly, the within-country C index will give slightly different values to the T index under options 1a2b3a4a unless sample sizes are very small. The other indices in Table 3 give exactly the same value as the T index with the options specified.

We illustrate the use of the formula for the two-country I index between the UK and Belgium (I = 0.17, SE = 0.05, see Table 2). From Table 1, and after ignoring companies not disclosing their accounting method, we have for the UK ($i = 4$), p_{k4} equal to 7/12, 0, 2/12, and 3/12 ($n_4 = 12$) for accounting methods k = 1 to 4 respectively, while for Belgium ($i = 5$) the corresponding proportions p_{k5} equal 2/19, 6/19, 8/19, and 3/19 ($n_5 = 19$). Using these as estimates for the π_{ki} we obtain the values for $\theta_{kl(34)}$ provided in Table 4, and summing these values gives a variance of 0.00248. The standard error of 0.05 reported in Table 2 is the square-root of this variance.

The formula for the variance of the between-country C index with more than two countries requires a table of $\theta_{kl(ij)}$, as well as a corresponding table of $\phi_{kl(ij)}$, for each pair of countries i and j. The $\phi_{kl(ij)}$ terms account for correlations between two-country I indices that have a country in common, such as the I index between countries 1 and 2 and the I index between countries 1 and 3.

The formulae for the bias and variance of the within-country C index and between-country C index in Table 3 are valid for any choice of the β_{ii} and β_{ij} respectively. The values $\beta_{ii} = n_i^2 / \sum_i n_i^2$ and $\beta_{ij} = n_i n_j / \sum_i \sum_{j \neq i} n_i n_j$ are specified in the formulae for the indices only because these correspond to option 1a (companies weighted equally) used by C indices. International studies that prefer, for example, to weight countries equally (option 1b) can use the formulae for bias and variance in Table 3 by specifying $\beta_{ii} = 1/N$ for the within index and $\beta_{ij} = 1/(N(N-1))$ for the between-country index.

4. The nine measurement practices of Herrmann and Thomas (1995)

Herrmann and Thomas (1995) examined the level of comparability in Belgium, Denmark, France, Germany, Ireland, the Netherlands, Portugal and the UK using data from the 1992–1993 annual reports of 217 companies. They used a modification of the I index by substituting values of 0.01 and 0.99 when the proportion of companies within a country using a particular method were 0 and 1 respectively. They argued this modification was necessary because 'the I index is sensitive to zero proportions' and that this 'potential sensitivity increases as the number of countries surveyed increases' (Herrmann and Thomas, 1995: 256). Taplin (2004) discussed problems with the I index for more than two countries and with this ad hoc adjustment. Table 1 reproduces the data for the measurement practice of inventory costing. Data for all nine measurement practices is available in Herrmann and Thomas (1995).

Table 5 contains the I and T index values for the data on the nine measurement practices in Herrmann and Thomas (1995). These T index

[2] A proof of this result is available from the author upon request.

Table 3
The formulae, bias and variance of the simple H, I and C indices that are special cases of the T index (with specified options). Standard errors equal the square-root of the variances.

	H index for one country[5] (1a2a3a4a)	Within-country C index (1a2b3a4a)[6,7]	I index with only two countries i and j (1a2c3a4a)[8]	Between-country C index with more than 2 countries (1a2c3a4a)[5]
Index formula	$H_i = \sum_k p_{ki}^2$	$C_w = \sum_i \beta_{ii} H_i$ where $\beta_{ii} = n_i^2 / \sum_i n_i^2$	$I_{ij} = \sum_{k=1}^{M} p_{ki} p_{kj}$	$C_b = \sum_i \sum_{j \neq i} \beta_{ij} \sum_k p_{ki} p_{kj}$ where $\beta_{ij} = n_i n_j / \sum_i \sum_{j \neq i} n_i n_j$
Bias	$(1 - H_i)/n_i$	$\sum_i \beta_{ii}(1 - H_i)/n_i$	0	0
Variance	$\sigma^2_{H_i}$	$\sum_i \beta_{ii}^2 \sigma^2_{H_i}$	$\sum_k \sum_l \theta_{kl(ij)}$	$2 \sum_i \sum_{j \neq i} \beta_{ij} \sum_k \sum_l (\beta_{ij} \theta_{kl(ij)} - \phi_{kl(ij)})$

k, l are dummy indicators for possible accounting methods and take values 1 to M
i, j, J are dummy indicators for possible countries and take values 1 to N

$\sum_i$ is a summation over all possible values of i from 1 to N

$\sum_{j \neq i}$ is a summation over all possible values of j from 1 to N except i

n_i is the number of sampled companies from country i.
p_{ki} is the proportion of companies from country i using method k.

$\sigma^2_{H_i} = \sum_k a_{k(i)} + \sum_k \sum_{l \neq k} b_{kl(i)} - \left(1 - (n-1) \sum_k \pi_{ki}^2\right)^2 / n_i^2$ is the variance of the H index for country i.

$$a_{k(i)} = (\pi_{ki} + 7(n_i - 1)\pi_{ki}^2 + 6(n_i - 1)(n_i - 2)\pi_{ki}^3 + (n_i - 1)(n_i - 2)(n_i - 3)\pi_{ki}^4)/n_i^3$$

$$b_{kl(i)} = ((n_i - 1)(n_i - 2)(n_i - 3)\pi_{ki}^2 \pi_{li}^2 + (n_i - 1)(n_i - 2)\pi_{ki}\pi_{li}(\pi_{ki} + \pi_{li}) + (n_i - 1)\pi_{ki}\pi_{li})/n_i^3$$

$\theta_{kl(ij)} = (1 - n_i - n_j)\pi_{ki}\pi_{kj}\pi_{li}\pi_{lj}/(n_i n_j)$ when $k \neq l$, and

$$\theta_{kk(ij)} = \left((n_i - 1)\pi_{ki}^2 + \pi_{ki}\right)\left((n_j - 1)\pi_{kj}^2 + \pi_{kj}\right)/(n_i n_j) - \pi_{ki}^2 \pi_{kj}^2$$

$\phi_{kl(ij)} = \pi_{ki}\pi_{kj} \sum_{J \neq i,j} (\pi_{li}\pi_{lJ}\beta_{iJ}/n_i + \pi_{lj}\pi_{lJ}\beta_{jJ}/n_j)$ when $k \neq l$, and

$$\phi_{kk(ij)} = \pi_{ki}\pi_{kj} \sum_{J \neq i,j} (\pi_{kJ}(\pi_{ki} - 1)\beta_{iJ}/n_i + \pi_{kJ}(\pi_{kj} - 1)\beta_{jJ}/n_j)$$

where the summations over J are for all possible values of from 1 to N except i and j.

[5] The H index and overall C index are slightly different due to differences between sampling with or without replacement, but these differences are negligible unless sample sizes are very small. We present results for a single country here since this is how the H and C indices are typically used. If a simple random sample from several countries is taken then these formulae should be used for the bias and variance of the T index under options 1a2a3a4a. If a stratified random sample is taken, where simple random samples of a pre-specified size are independently taken from each country, the general formula for the bias and variance of the T index should be used.

[6] The T (1a2b3a4a) and within-country C indices are slightly different due to differences between sampling with or without replacement, but these differences are negligible unless sample sizes are very small.

[7] These formulae for the bias and variance of the within-country C index and the between-country C index hold for other choices of weightings β_{ij}. For example, if countries are weighted equally (option 1b) rather than companies weighted equally (1a) assumed by the C indices, then $\beta_{ii} = 1/N$ for the within index and $\beta_{ij} = 1/(N(N-1))$ for the between-country index. The provided formulae for the bias and variance of these indices still hold with these weighting of companies/countries.

[8] This equals the between-country C index since there are two countries.

Table 4
Values of $\theta_{kl(34)}$ used to compute the standard error for the two-country I index (or between-country C index) between the UK and Belgium for inventory costing in the Herrmann and Thomas (1995) data after removing non-disclosing companies.

k	*l*			
	1	*2*	*3*	*4*
1	0.002012	0.000000	–0.000243	–0.000593
2	0.000000	0.000000	0.000000	0.000000
3	–0.000243	0.000000	0.002557	–0.000678
4	–0.000593	0.000000	–0.000678	0.000936

The sum of these 16 values gives the variance of the *I* index equal to 0.00248. The standard error of 0.05 reported in Table 2 is the square-root of 0.00248.
Values of $\theta_{kl(34)}$ in this table are calculated using the formula for $\theta_{kl(ij)}$ beneath Table 3, with $i = 4$ (the UK) and $j = 5$ (Belgium) and the proportions estimated from the sample data in Table 1 as follows: p_{k4} equal to 7/12, 0, 2/12, and 3/12 ($n_4 = 12$) and p_{k5} equal 2/19, 6/19, 8/19, and 3/19 ($n_5 = 19$) for $k = 1$ to 4 respectively.

values were calculated using options 1b2c3a4a, so as with the *I* index countries are weighted equally, comparisons are between different countries only, multiple accounting policies are not allowed and companies not disclosing their method are removed from the sample. Thus for measurement practice 6, $T = 0.27$ equals the simple average of the off-diagonal entries in Table 2. Total sample sizes after removing non-disclosing companies and the standard errors and 95% confidence intervals for the *T* indices are also presented.

From Table 5 we see that the *I* index and *T* index values are generally close, but differ by more than 0.10 for measurement practices 1, 3 and 4. The data for all of these contain zero proportions and hence are influenced by the arbitrary values of 0.01 and 0.99 substituted by Herrmann and Thomas (1995). For measurement practices 1 and 4 there are sufficient zero proportions in the data for the unmodified *I* index to equal 0. The *I* index in these circumstances is unstable, depending on whether the particular sample has proportions of zero or not and on the value of the arbitrary values of 0.01 and 0.99 substituted.

The standard errors for the *T* index are generally small and indicate that the values of the *T* index calculated from this sample are accurate estimates. This is illustrated in Figure 2 where ± 2 standard error bars (representing 95% confidence intervals for the population index values) are presented graphically. This provides statistically significant evidence that the level of comparability for measurement practice 6 (Inventory costing) is lower than for any of the other eight practices since the 95% confidence intervals are far from overlapping. The evidence for measurement practice 7 (Foreign currency translation of assets and liabilities) having the highest level of comparability is much weaker. Although in the sample data the index value of 0.87 for practice 7 is the highest, its confidence interval from 0.81 to 0.92 overlaps considerably with the confidence interval from 0.69 to 0.90 for practice 9 (Treatment of translation differences).

From Table 5 and Figure 2 we see that measurement practice 9 (Treatment of translation differences) has a substantially larger standard error of 0.053 compared to the other practices. This is partially explained by the small sample size resulting from the high level of non-disclosure for this practice. Table 5 reveals, however, other measurement practices such as 4 (Research and development) with both a smaller sample size and standard error.

The reason for the high standard error of 0.053 for practice 9 is because 15 of the 20 sampled companies from Portugal did not disclose their accounting method, resulting in an effective sample size of only 5 for Portugal. The higher standard error results from the fact that any comparisons between Portugal and another country cannot be made with statistical confidence. If companies had been weighted equally (option 1a under the *T* index), then the standard error is reduced from 0.053 to 0.033, and similar to the standard errors for most of the other practices. Although comparisons with a Portugese company are prone to statistical inaccuracy, these comparisons are given little weight under option 1a because there is a small sample of companies from Portugal.

We do not suggest that the lower standard error under option 1a means that preference should be given to option 1a rather than option 1b when using

Table 5
Total sample sizes *n* for disclosing companies (summed over all countries), I indices and T (options 1b2c3a4a) indices (with standard errors and 95% confidence intervals) for the Herrmann and Thomas (1995) data.

Measurement practice	*n*	*I*	*T*	*SE*	*95%*	*CI*
1. Fixed asset valuation	217	0.29	0.47	0.006	0.46	0.48
2. Depreciation	217	0.62	0.68	0.023	0.64	0.72
3. Goodwill	187	0.25	0.45	0.005	0.44	0.46
4. Research and development	109	0.41	0.58	0.026	0.53	0.63
5. Inventory valuation	217	0.79	0.71	0.032	0.65	0.77
6. Inventory costing	124	0.23	0.27	0.017	0.24	0.30
7. Foreign currency translation of assets and liabilities	188	0.90	0.87	0.028	0.81	0.92
8. Foreign currency translation of revenues and expenses	184	0.64	0.60	0.032	0.54	0.67
9. Treatment of translation differences	179	0.85	0.79	0.053	0.69	0.90

the T index. As described in Taplin (2004), the choice of options under the unified framework of the T index should be tailored to the specific research question being addressed. Rather, the example in the previous paragraph highlights the fact that it is more important to have higher sample sizes in *each* country when equal weighting is given to each country than when companies are given equal weighting. The general principle is as follows: to obtain a more accurate value for the T index (that is, a lower standard error), the sample size should be higher for countries that are given higher weight in the T index. For option 1b this suggests approximately equal sample sizes in each country.

5. A comparison between fairness and legalistic countries

Herrmann and Thomas (1995) also compared the level of comparability between the fairness countries (Denmark, Ireland, the Netherlands and the UK) to the level of comparability between legalistic countries (Belgium, France, Germany and Portugal). They concluded on p. 264 that 'The bicountry and four-country I indices reveal that fairness oriented countries are more harmonised than legalistic ones'. They made no attempt to examine the statistical significance of these differences. Taplin (2003) reported this statistical comparison, however only after using the H index instead of the I index. In Taplin's (2004) unified framework of the T index this required companies to be weighted equally (option 1a) instead of countries being weighted equally (option 1b) and an overall international perspective (option 2a) rather than a between-country perspective (option 2c).

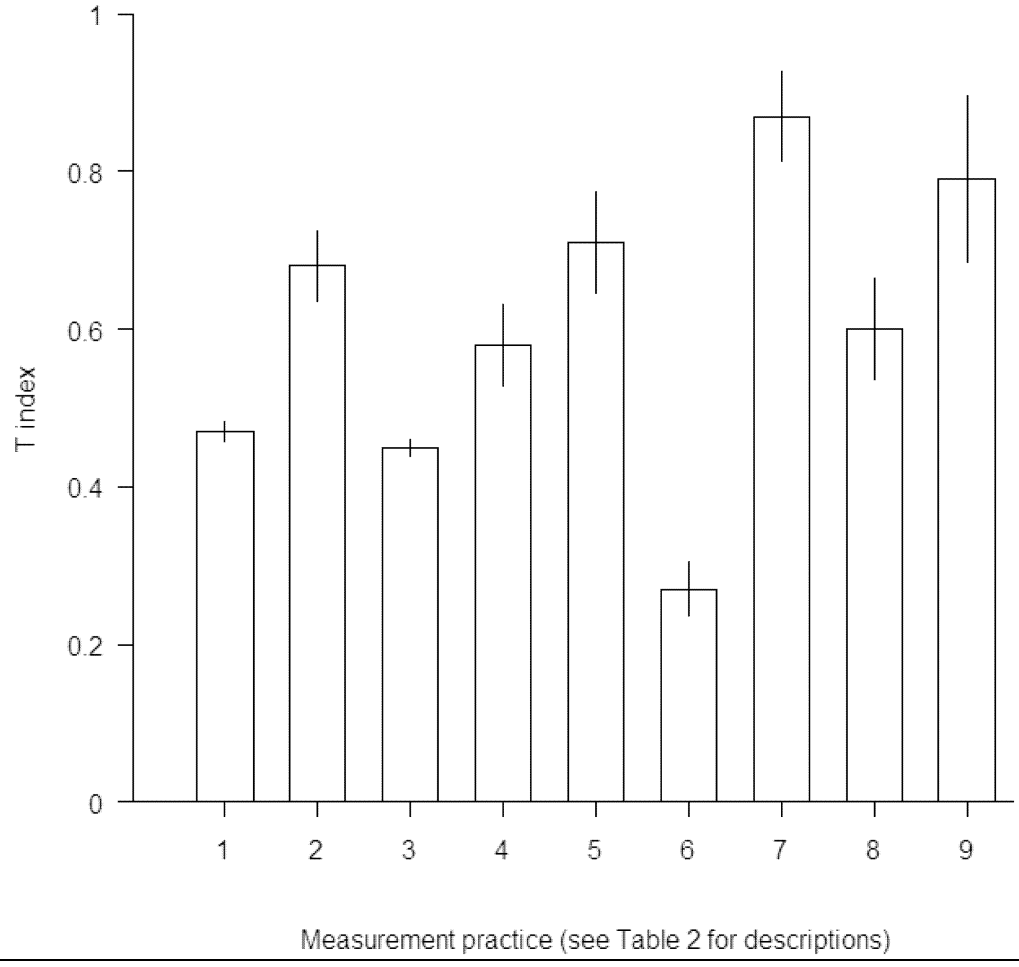
Here we present the statistical comparison between the fairness and legalistic countries using options 1(b) and 2(c) of the T index. For comparison purposes with the results in Taplin (2003), we begin by retaining the options 3(a) and 4(a) so, as with the H Index in Taplin (2003), non-disclosing companies are removed prior to analysis.

Table 6 contains the values for the T index for the fairness and legalistic countries together with their standard errors. These are the four-country indices that are most comparable with the I index used by Herrmann and Thomas (1995), but, avoids problems associated with the various forms of the I index for more than two countries. Table 6 also contains the difference in T index values (Fairness countries T index minus legalistic countries T index), the standard error of this difference, and standardised score Z and P-value when testing the null hypothesis of no difference in T index values. Note that the standard error for the difference in T index values equals the square-root of the sum of the squares of the two standard errors. For example, for the first measurement practice (Fixed asset valuation), the standard error for the difference in T index values equals $\sqrt{0.017^2 + 0.011^2} = 0.020$.

As reported by Herrmann and Thomas (1995) the level of comparability is higher in the fairness countries for seven of the nine measurement practices. Measurement practice 3 (Goodwill) is one of the exceptions, but Table 6 shows that this difference of –0.06 is not statistically significant (P = 0.396). Thus this data is consistent with no difference in the level of comparability in fairness and legalistic countries, and indeed is consistent with either fairness or legalistic countries having the higher level of comparability.

The other exception is measurement practice 8 (Foreign currency translation of revenues and expenses). In this case, not only is the level of comparability in the fairness countries smaller than the level in the legalistic countries (T= 0.48 compared to T = 0.78), but this difference is highly significant statistically (P = 0.000). Although the overall trend

Figure 2
T index (with ±2 standard error bars) for the Herrmann and Thomas (1995) data



reported by Hermann and Thomas (1995) that fairness countries are more comparable is valid, this highly significant trend in the opposite direction for foreign currency translation of revenues and expenses may deserve further examination.

Furthermore, for two of the seven measurement practices where the level of comparability is higher in the fairness countries compared to the legalistic countries, the difference in the level of comparability is not statistically significant (P = 0.174 and P = 0.272 for practices 1 and 5). Thus there is statistically significant evidence that comparability is higher for fairness rather than legalistic countries in only five of the nine measurement practices. The addition of this statistical rigour adds clarity to our interpretation of index values.

6. The effect of non-disclosure on the standard error

Pierce and Weetman (2002) warned that interpretation of index values was problematic when the non-disclosure level was high. Unlike the *I* index employed by Herrmann and Thomas (1995) and previous *H* and *C* indices, their adjusted *C* index does not remove all non-disclosing companies. Their analysis was, however, restricted by the lack of statistical inference techniques presented in this paper. We therefore repeat our comparison of the fairness and legalistic countries assuming companies not disclosing their accounting method are not comparable to all other companies.[3] This can be achieved by using option (4c) instead of (4a) under the *T* index framework. Results appear in Table 7 using the same format as in Table 6. For measurement practices 1, 2 and 5 all companies in the sample data disclosed their accounting method. In these cases results under option (4c) are identical to

[3] For brevity we do not present results assuming non-disclosing companies are comparable with all other companies or, as Pierce and Weetman (2002) suggest, determining which companies are applicable and which are not-applicable non-disclosures.

Table 6
A statistical comparison of the T indices (option 1b2c3a4a) for the fairness and legalistic counties and the Herrmann and Thomas (1995) data.

Measurement practice	*Fairness countries*		*Legalistic countries*		*Difference*			
	T	*SE*	*T*	*SE*	*T*	*SE*	*Z*	*P-value*
1. Fixed asset valuation	0.48	0.017	0.45	0.011	0.03	0.020	1.36	0.174
2. Depreciation	0.95	0.030	0.47	0.022	0.48	0.037	12.85	0.000***
3. Goodwill	0.79	0.052	0.84	0.043	–0.06	0.068	–0.85	0.396
4. Research and development	0.83	0.064	0.38	0.021	0.45	0.068	6.60	0.000***
5. Inventory valuation	0.73	0.048	0.65	0.053	0.08	0.071	1.10	0.272
6. Inventory costing	0.63	0.084	0.25	0.024	0.38	0.088	4.34	0.000***
7. Foreign currency translation of assets and liabilities	0.98	0.018	0.76	0.050	0.22	0.053	4.21	0.000***
8. Foreign currency translation of revenues and expenses	0.48	0.025	0.78	0.072	–0.30	0.076	–3.90	0.000***
9. Treatment of translation differences	0.92	0.042	0.67	0.091	0.25	0.100	2.48	0.013*

*, ** and *** denotes statistical significance at the 0.05, 0.01 and 0.001 level respectively (two-tailed tests).

results under option (4a) and so are not repeated in Table 7.

As expected, the change from option (4a) to (4c), whereby non-disclosing companies are considered non-comparable with all other companies, results in a lower level of comparability when non-disclosure exists. All *T* index values in Table 7 are smaller than the corresponding value in Table 6. Also as expected, these decreases are greatest where the level of non-disclosure is highest. For example, measurement practices 4 and 6 for the fairness countries have both the highest non-disclosure rate (with respectively 108 and 93 non-disclosing companies out of 217) and the highest reduction in *T* index values, from 0.83 and 0.63 (Table 6) to 0.25 and 0.09 (Table 7) respectively.

Smaller sample sizes are typically associated with larger standard errors. However, removing non-disclosing companies does not necessarily increase standard errors. Exactly half of the 12 standard errors for *T* indices in Table 7 (with non-disclosing companies included) are smaller than the corresponding standard errors in Table 6 (where non-disclosing companies are removed). For measurement practice 7 in the fairness countries, the standard error triples from 0.018 to 0.055 when non-disclosing companies are included. This is because the very high value of 0.98 for the *T* index under option (4a) is close to the boundary of 1 and this constrains the size of its standard error. This is discussed in Section 8 where we consider the shape of the sampling distribution of the *T* index.

The treatment of non-disclosure can have a large effect on the comparison between the level of comparability within the fairness and legalistic countries. Tables 3 and 4 both indicate a significant difference in the level of comparability within fairness and legalistic countries for measurement practice 6 (P = 0.000 and P = 0.023 respectively). However, the sign of the difference is not the same: under option (4a) where non-disclosing companies are removed the fairness countries have the higher level of comparability but under option (4c) where non-disclosing companies are considered non-comparable the legalistic countries have the higher level of comparability. In this case conclusions concerning which group of countries is more comparable depends significantly on the treatment of non-disclosure. In contrast, we note that while the difference in *T* index values in Tables 3 and 4 for measurement practice 3 have different signs, neither is significantly different to zero (P = 0.396 and P = 0.314 respectively). No ambiguity in conclusions arises for measurement practice 3 because there is insignificant evidence of any difference in the level of comparability for fairness compared to legalistic countries. Although p-values in Tables 3 and 4 for the other measurement practices do change, conclusions remain the same if they are based on the conventional signifi-

Table 7
A statistical comparison of the T indices (option 1b2c3a4c) for the fairness and legalistic counties and the Herrmann and Thomas (1995) data.

Measurement practice	*Fairness countries*		*Legalistic countries*		*Difference*			
	T	*SE*	*T*	*SE*	*T*	*SE*	*Z*	*P-value*
1. Fixed asset valuation								
2. Depreciation								
3. Goodwill	0.61	0.057	0.52	0.062	0.08	0.084	1.01	0.314
4. Research and development	0.25	0.046	0.06	0.015	0.19	0.048	3.85	0.000***
5. Inventory valuation								
6. Inventory costing	0.09	0.025	0.16	0.022	–0.07	0.033	–2.27	0.023*
7. Foreign currency translation of assets and liabilities	0.74	0.055	0.49	0.061	0.24	0.082	2.98	0.003**
8. Foreign currency translation of revenues and expenses	0.35	0.033	0.52	0.057	–0.18	0.066	–2.66	0.008**
9. Treatment of translation differences	0.65	0.057	0.40	0.049	0.25	0.075	3.31	0.001***

*, ** and *** denotes statistical significance at the 0.05, 0.01 and 0.001 level respectively (two-tailed tests).

cance level of 5% and on the direction of the difference.

7. The special case of the two-country I index

As shown in Section 2 the two-country *I* index, or the equivalent between country *C* index for two countries, is important because it is a basic ingredient in the calculation of the *T* index. We investigate the statistical properties of this special case of the *T* index in this section because some studies will compare only two countries. This will also provide additional insights into the accuracy with which *T* index values can be estimated (and later in Section 8 the extent to which the sampling distribution deviates from normality) when sample sizes are small. Table 2 contains the standard errors (in parentheses) for each of the two-country *I* indices for measurement practice 6 (Inventory costing) and Section 3.2 illustrates the calculation of these standard errors.

Fairness countries (listed first in Table 2) have higher values for the two-country *I* index (ranging from 0.49 to 0.82) than legalistic countries (ranging from 0.10 to 0.34) or between a fairness and legalistic country (ranging from 0.00 to 0.26). They also have higher standard errors (ranging from 0.11 to 0.14 compared to 0.05 to 0.08 and 0.00 to 0.08 respectively), reflecting the lower level of accuracy with which we have estimated the degree of comparability between fairness countries. This is largely due to the smaller sample sizes for the fairness countries, which result in part from the higher level of non-disclosure in these countries. The lowest level of comparability between two fairness countries is 0.49 (SE = 0.11) between Denmark and the UK. This is not significantly higher than the comparability between some pairs of legalistic countries. For example, the level of comparability between France and Portugal is 0.34 (SE = 0.05).

Finally, we note for measurement practice 6 in Table 5 the standard error of 0.017 is considerably lower than 0.071, the mean of the off-diagonal standard errors in Table 2. Recall (see Section 2) the corresponding value of $T = 0.27$ in Table 5 is the mean of the off-diagonal index values in Table 2. The lower standard error in Table 5 is a direct result of the *T* index using data from all eight countries while each of the indices in Table 2 uses data from only two countries: indices based on larger samples are expected to have similar values, on average, but with smaller standard errors.

This smaller standard error for an index based on several countries compared to the standard error for an index based on two countries has implications when designing studies. The recommendation in Taplin (2004) to examine the two-country *I* indices whenever calculating a *T* index with several countries is still relevant, however if research questions relate to comparisons between pairs of

countries as well as between all countries then larger sample sizes may be required. We discuss the issue of required samples sizes in Section 9.

8. The sampling distribution for the T index

This paper has presented formulae for the standard error of the T index and applied them to several examples. The formulae are exact under independent random sampling of companies from countries without any additional assumptions, but the confidence intervals and p-values derived from these standard errors also assume that the sampling distribution of the T index is normal. For example, 95% confidence intervals were constructed with the estimated T index plus or minus 1.96 standard errors and Z scores were converted into p-values using standard normal tables. In practice, normality is often a good approximation for this sampling distribution, especially in large sample sizes n_i or when there is a large number of countries, due to Central Limit Theorem results. First, when the sample size n_i is large the corresponding p_{ki} for that country will be approximately normally distributed. Second, since the T index is a weighted average of random variables, when this average is over a larger number of terms the sampling distribution is likely to be closer to normal.

Central Limit Theorem results, however, provide limits as sample sizes tend to infinity and are therefore of theoretical interest. Instead, we provide some examples of the sampling distribution for the T index so the accuracy of the normal distribution can be evaluated in more realistic finite samples. We do so by providing histograms of one million T index values generated from one million simulated samples from known populations. Not only do these simulations allow a comparison with the normal distribution, the interval from the 0.025 percentile to the 0.975 percentile of these distributions provides an exact 95% interval to compare with the approximate plus or minus 1.96 standard error approximation based on normality. Since the approximation is extremely accurate in most examples covered in this paper, we concentrate on cases where the approximation is weakest. Although two decimal places are generally sufficient when using a confidence interval to assess the accuracy of an estimate, in this section we quote intervals to three decimal places to allow closer scrutiny of the accuracy of this normality approximation.

A value of $T = 0.98$ (and standard error of 0.018) was reported in Table 6 for fairness countries and measurement practice 7. In this case, it is clear that the sampling distribution can not be normal because the upper bound for a legitimate T index value is 1, only just over one standard error from the estimate. Furthermore, a 95% confidence interval calculated with the plus or minus 1.96 standard error rule will be non-sensible in this case since it will include values beyond the theoretical boundaries of 0 and 1. Of the 54 T index values presented in Tables 2, 3 and 4 there are three occasions where these approximate confidence intervals for the T index using the plus or minus 1.96 standard errors rule result in intervals extending past either 0 or 1. These are for fairness countries with measurement practices 7, 2, and 9 (where T index values are 1.0, 1.8 and 1.9 standard errors from the nearest boundary). We also include measurement practice 4 for the fairness countries since its T index value is 2.6 standard errors from the boundary of 1. No other T index is within 3 standard errors of a boundary of 0 or 1.

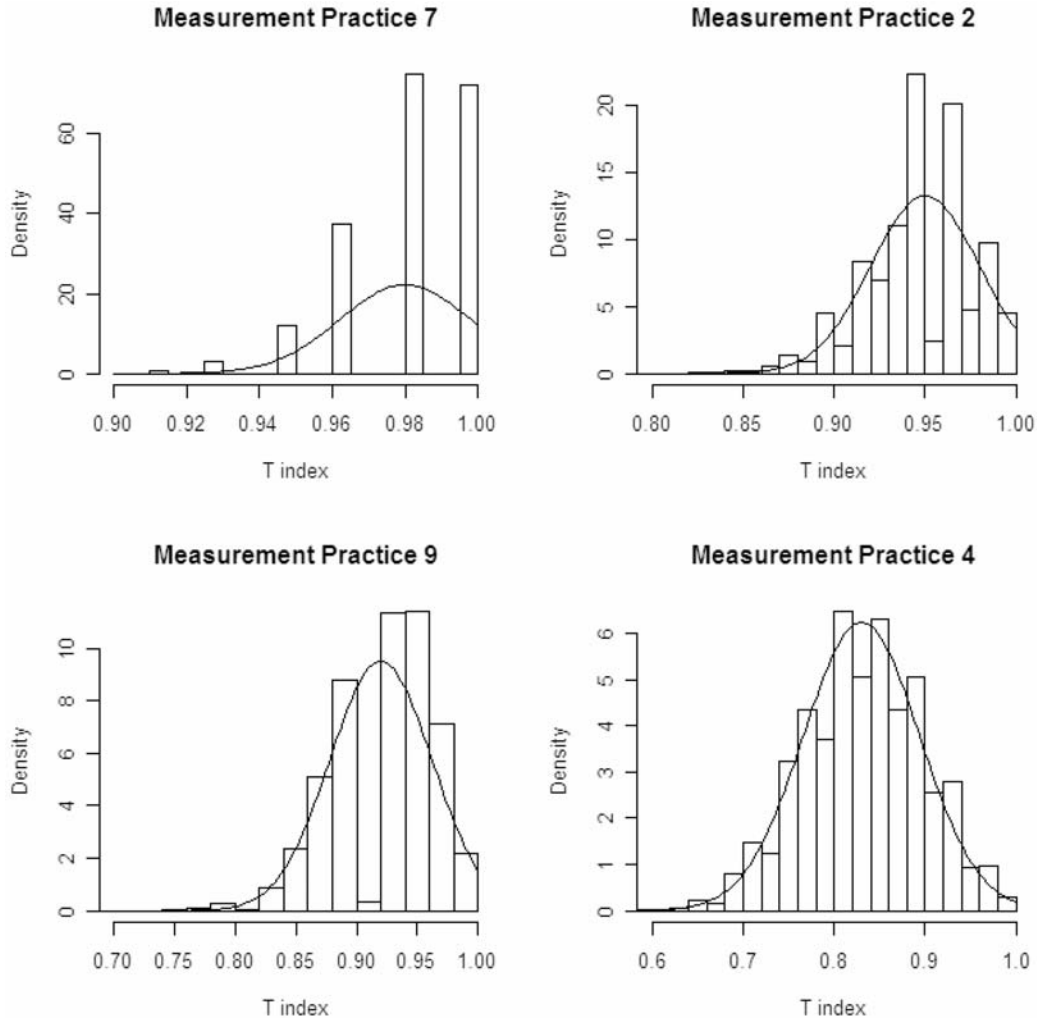
Figure 3 presents the estimated sampling distribution from the results of simulating one million samples from each of these populations. That is, these histograms describe the probability of obtaining different values for the T index when randomly sampling companies. Each distribution has the normal distribution superimposed that uses the theoretical standard error given in Section 3.2. Note that in each of these cases the bias is zero since the indices use a between country international perspective (option 2c).

Measurement practice 7 (top left of Figure 3) shows a sampling distribution that is clearly not normal, as expected since the mean of the distribution is only 1.0 standard errors from the boundary of 1. It is highly skewed and discrete in nature with only a few index values possible. Indeed, this sampling distribution only contains the possible values of 1 and multiples of 0.018 less than 1 (1, 0.982, 0.964,$\cdots$) corresponding to none, one, two,$\cdots$sampled companies using the current/historical method rather than the historical method. Only one of the 99 companies from the fairness countries in this data used the current/historical method.

Despite this clear non-normality for measurement practice 7, the approximate 95% interval calculated using the plus or minus 1.96 standard errors rule is accurate. In this case, the approximation gives the interval from 0.948 to 1.017. The 95% interval calculated using the 0.025 and 0.975 percentiles from the million simulations yields an interval from 0.946 to 1.

Measurement practices 2, 9 and 4 show progressively lower degrees of non-normality with the means of the sampling distributions being 1.8, 1.9 and 2.6 standard errors from a boundary. Their

Figure 3
Sampling distributions based on 1,000,000 samples for the T index corresponding to measurement practices 7, 2, 9 and 4 in Table 6 for fairness countries. Normal approximations based on theoretical standard errors are also shown.



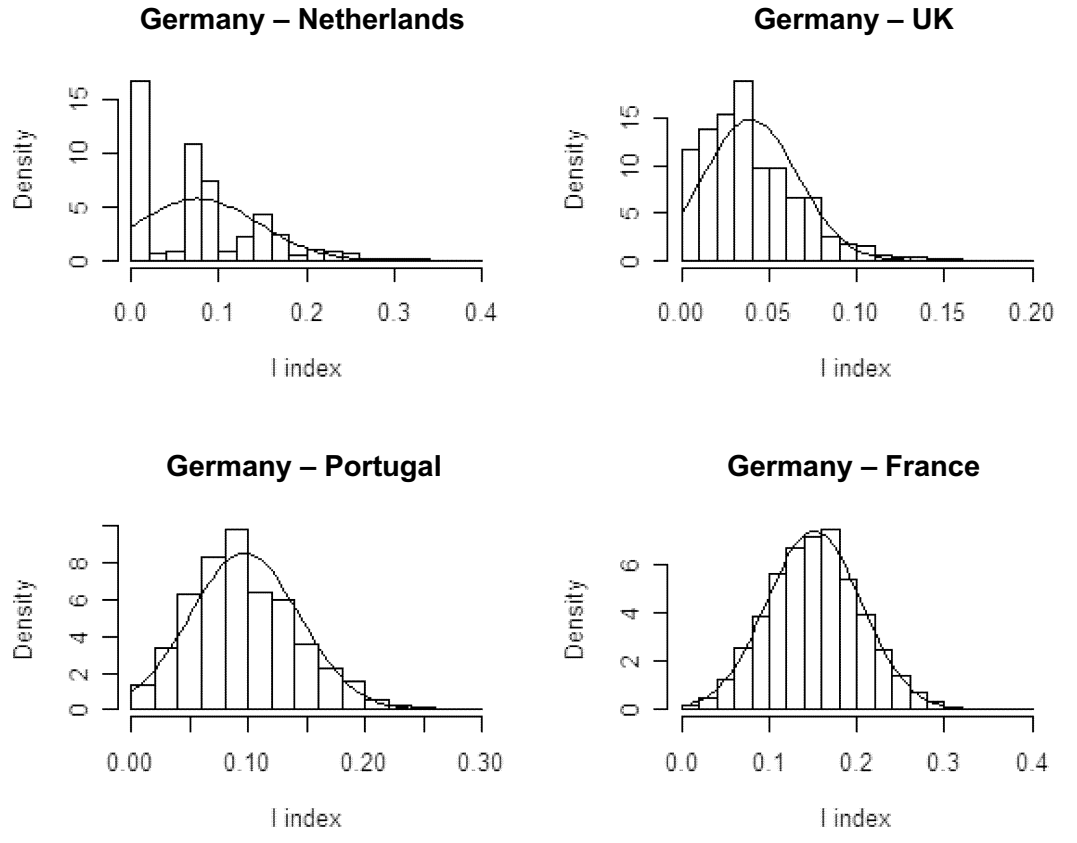
approximate intervals are 0.889 to 1.005, 0.839 to 1.004 and 0.705 to 0.957 respectively. The intervals based on the percentiles are 0.882 to 1, 0.823 to 1 and 0.694 to 0.962 respectively. In each case the approximate intervals are accurate within 0.02 and usually accurate within 0.01. Hence the approximations are extremely accurate, especially compared to the inaccuracy in the estimated T index values summarised by the length of the intervals.

All of the other 50 T index values presented in Tables 2, 3 and 4 are more than three standard errors from the boundaries of 0 and 1 and the approximate 95% confidence intervals formed by taking the estimated T index value plus or minus 1.96 standard errors are very accurate. Practical experience suggests that this approximation is extremely accurate when the T index is more than three standard errors from both boundaries of 0 and 1. Furthermore, it is likely to be accurate to within 0.01 if the T index is more than two standard errors from both boundaries of 0 and 1, and often accurate even when this is not the case.

Hence we conclude that statistical inference performed as if the sampling distribution of the T index is normal is likely to be sufficiently accurate for most purposes unless possibly when the estimated index value is within two standard errors of a boundary. This is likely to occur when either there is a very high or low level of comparability resulting in a T index value that is close to either 0 or 1, or the standard error is very high reflecting an inaccurate estimate of the population T_p index value due to samples sizes that are too small.

Normality is less likely to be a valid approxima-

Figure 4
Sampling distributions based on 1,000,000 samples for the I index corresponding to measurement practice 6 and comparisons between Germany and either the Netherlands, the UK, Portugal or France. Normal approximations based on theoretical standard errors are also shown.



tion for the two-country I indices for Inventory costing in Section 7. This is because these T indices are calculated from a sum involving fewer terms and the standard errors tend to be higher. Figure 4 displays the sampling distribution for some of the two-country I indices in Table 2. The Netherlands, the UK, Portugal and France were chosen for illustration purposes because the index values between these countries and Germany are close to a boundary of 0 or 1 (respectively 1.1, 1.5, 2.1 and 2.8 standard errors away).

From Figure 4 it is apparent that normality is a poor approximation when the value for the T index is close to a boundary of 0 or 1 (compared to the size of the standard error). For the Netherlands and the UK, approximate 95% confidence intervals using the value of the T index plus or minus 1.96 standard errors result in intervals including negative values (–0.059 to 0.212 and –0.013 to 0.092 respectively). The intervals from the 0.025 and 0.975 percentiles of the simulated distribution are 0.000 to 0.244 and 0.000 to 0.105 respectively. For Portugal, the sampling distribution is slightly skewed but close to normal. The approximate 95% confidence interval of 0.005 to 0.189 is close to the interval of 0.019 to 0.199 from the simulation. In this case the value for the I index is just over two standard errors from the closest boundary. Finally, for the I index comparing Germany and France the I index is 2.8 standard errors from the boundary and the sampling distribution is very close to normal. The approximate 95% confidence interval 0.046 to 0.256 and interval of 0.048 to 0.259 from the simulation are very close.

Once again these approximations are accurate, especially if the presence of intervals extending beyond the boundaries of 0 and 1 are ignored and the accuracy of the approximate intervals are compared to the length of the intervals.

9. Sample size determination

The formula for the standard error of the T index can be used to determine the necessary sample sizes for each country in order to achieve a given level of precision for the T index. We now illustrate this procedure for a simple example.

Table 8
Anticipated proportions π_{ki} for the proportion of companies within country *i* using accounting method *k*

Accounting method	*Country* 1	2	3	4
A	0.2	0.1	0.2	0.3
B	0.8	0.4	0.6	0.6
ND	0	0.5	0.2	0.1

Suppose we are planning a study involving four countries and we anticipate three accounting methods, labeled A, B and ND for non-disclosure. First we select the T index we desire for our study. Companies do not provide enough information in their accounts to enable comparability with both a company using method A and another company using method B so multiple accounting policies are not possible (option 3a). Furthermore, we consider non-disclosure to be not comparable (option 4c). We wish to make between-country comparisons (option 2c) and give each country equal weight (option 1b).

Second, we need to anticipate the proportion of companies within each country using each method. These are the population proportions π_{ki}. At first this appears strange in that we need to anticipate the characteristics of the population because if these were known we could calculate the exact value of the population T indices, but this is true of all sample size calculations. Suppose these proportions π_{ki} are given by the values in Table 8. For example, in the first country 20% of companies use method A, 80% use method B and all companies disclose their accounting method.

The value of the T index using options 1b2c3a4c for this population is $T_p = 0.39$. This is the value we are trying to estimate from our samples of companies. Now suppose our sample size is $n_i = 10$ for each of the four countries. Then from the equation for the standard error of the sample T index in Section 3.2 the standard error for the T index is 0.078. Since the value of the index is approximately five standard errors from the nearest boundary we can be confident that the sampling distribution for the T index is very close to normal (simulations confirm this). Hence we can be 95% confident that the value of the T index in our sample will be within $1.96 \times 0.078 = 0.153$ of the population value $T_p = 0.39$. If the level of accuracy $\pm$ 0.153 for an estimated value for the T index is insufficient then sample sizes will need to be increased.

Now suppose we wish to test the null hypothesis that the level of comparability between countries 1 and 2 is equal to the level of comparability between countries 3 and 4. The value of the T index applied to the population of countries 1 and 2 is $T_{p1,2} = 0.34$ and the value when applied to countries 3 and 4 is $T_{p3,4} = 0.42$. The difference in population T index values is $\Delta T_p = T_{p1,2} - T_{p3,4} = -0.08$ and thus the null hypothesis is false. It remains to calculate the probability of rejecting this null hypothesis for a particular sample size.

For sample sizes of $n_i = 10$ for each country, the standard errors for the sample T indices are 0.129 for countries 1 and 2 and 0.111 for countries 3 and 4. Hence the standard error of the difference in sample T index values is $\sqrt{0.129^2 + 0.111^2} = 0.170$. This represents a high level of imprecision in our estimated difference in comparability when the actual difference is only 0.08. Larger sample sizes will be required to have a high chance of rejecting this null hypothesis.

More formally, our test statistic is $Z = \frac{T_{1,2} - T_{3,4}}{SE}$ where $T_{1,2}$ and $T_{3,4}$ are the sample T indices for countries 1 and 2 and for countries 3 and 4 respectively, and SE is the standard error of the difference $T_{1,2} - T_{3,4}$. We reject the null hypothesis at the 5% significance level when $Z < -1.96$ or $Z > 1.96$ (two-sided test) since under the null hypothesis Z has a standard normal distribution. Under the alternative hypothesis where $\Delta T_p = T_{p1,2} - T_{p3,4}$ is non-zero, $\frac{T_{1,2} - T_{3,4} - \Delta T_p}{SE}$ has a standard normal distribution and hence the probability of rejecting the null hypothesis in favour of a two-sided alternative hypothesis is:

$$\text{Power} = P(Z < -1.96) + P(Z > 1.96)$$

which by definition of the test statistic Z,

$$= P\left(\frac{T_{1,2} - T_{3,4}}{SE} > 1.96\right) + P\left(\frac{T_{1,2} - T_{3,4}}{SE} < -1.96\right)$$

$$= P\left(\frac{T_{1,2} - T_{3,4} - \Delta T_p}{SE} > 1.96 - \frac{\Delta T_p}{SE}\right)$$

$$+ P\left(\frac{T_{1,2} - T_{3,4} - \Delta T_p}{SE} < -1.96 - \frac{\Delta T_p}{SE}\right)$$

and since $\frac{T_{1,2} - T_{3,4} - \Delta T_p}{SE}$ has a standard normal distribution,

$$= 1 - \Phi\left(1.96 - \frac{\Delta T_p}{SE}\right) + \Phi\left(-1.96 - \frac{\Delta T_p}{SE}\right)$$

where $\Phi()$ denotes the cumulative distribution function for the standard normal distribution. For the above example with sample sizes of 10 from

Table 9
Standard errors for the four-country index T (options 1b2c3a4c), the two-country indices $T_{1,2}$ and $T_{3,4}$, the difference $T_{1,2} - T_{3,4}$, and the power of a test for a difference in the two-country indices, for a variety of sample sizes n_i per country.

Sample sizes n_i	*Standard error for*				*Power testing*
	T	$T_{1,2}$	$T_{3,4}$	$T_{1,2} - T_{3,4}$	$T_{p1,2} \neq T_{p3,4}$
10	0.078	0.129	0.111	0.170	0.076
20	0.055	0.090	0.076	0.118	0.104
30	0.045	0.073	0.062	0.096	0.133
50	0.035	0.057	0.047	0.074	0.192
100	0.024	0.040	0.033	0.052	0.337
200	0.017	0.028	0.023	0.037	0.588
500	0.011	0.018	0.015	0.023	0.932

each country, $\Delta T_p = -0.08$ and $SE = 0.170$, so $\frac{\Delta T_p}{SE} = -0.47$ and the power is $1 - \Phi(2.43) + \Phi(-1.49) = 0.076$, only marginally higher than the 5% probability of rejecting the null hypothesis if the null hypothesis was true.

While the normality assumption is reasonable in this case (the T index values are expected to be approximately 0.34/0.129 = 2.6 and 0.42/0.111 = 3.8 standard errors from the nearest boundary), this calculation is approximate because it ignores any differences between the estimated standard errors from the sample data and the standard errors calculated from the population proportions π_{ki}. These differences will be small in large samples. More importantly, any sample size calculation is always approximate since anticipated characteristics of the population (the π_{ki} in this case) will generally lead to greater inaccuracies in power calculations.

Table 9 summaries the relationship between sample size and the standard error for the four-country index T, the two-country T indices $T_{1,2}$ and $T_{3,4}$, the difference $T_{1,2} - T_{3,4}$, and the power of a test for a difference in the two-country T indices. In all cases the sample sizes are assumed equal, so the total sample size for the study is four times the value n_i given in Table 9.

The power of the test of the alternative hypothesis that $T_{p1,2} \neq T_{p3,4}$ is low for moderate sample sizes. From Table 9, the number of companies sampled from each country must be at least about 200 (total sample size of 800 companies) before there is at least a 50% chance of obtaining significant evidence for a difference in the level of comparability. It is unlikely that resources will be available for such a large study. From this we conclude that a study aiming to prove that the level of comparability between countries 1 and 2 is different from the level of comparability between countries 3 and 4 is not worth pursuing because, for sample sizes that are realistic in practice, the probability of achieving this aim is small.

The above is for illustration purposes and should not be taken to mean that all studies involving the T index require large sample sizes. For example, from Table 9 we see that samples of 30 companies from each of the four countries will result an estimated T index that is accurate within 9% (1.96 times the standard error of 0.045). Furthermore, the reason for the low power when testing $T_{p1,2} \neq T_{p3,4}$ was because the calculations assumed the levels of comparability $T_{p1,2}$ and $T_{p3,4}$ are very close, differing by only 0.08. Studies will not generally seek to detect such a small difference and smaller sample sizes will suffice to detect larger differences in comparability. Finally, in Sections 5 and 6 we saw that the sample sizes in the Herrmann and Thomas (1995) data were sufficient to detect some differences between comparability within fairness and legalistic countries.

10. Discussion

The T index is a flexible framework that enables researchers to select or design a particular index from many individual indices to suit their particular research question and the characteristics of their data. This paper enhances the T index by providing formulae for the bias and standard error of the T index and illustrates how the standard error can be used to compute standard statistical quantities such as confidence intervals and p-values for hypothesis tests. These techniques will add substantially to the value of research quantifying the level of comparability with the T index. Indeed, it is argued that every time a T index is reported in the literature a corresponding standard error (or confidence interval) should also be included to inform the reader how accurate the level of comparability has been measured with the sample data available.

As well as providing measures of accuracy of estimated T index values from samples, the methods illustrated in this paper enable the statistical testing of formal hypotheses. Examples of hypotheses include 'the level of comparability for accounting practice A is greater than the level of comparability for accounting practice B' or 'the level of comparability in countries A–C is greater than the level of comparability in countries D–G'. The results of this paper make this possible using any index from within the versatile T index framework.

The research agenda of Nobes (2006) proposed eight hypotheses and several sub-hypotheses worthy of future research in international accounting. From his detailed examination of motives and opportunities he concluded differences in IFRS practice internationally will persevere. 'The implications for users of IFRS financial statements are that international comparability may have increased but that large differences are likely to remain' Nobes (2006: 244). Importantly, the comparability of company accounts is the ultimate issue. While other issues such as compliance with IFRS are important because they can impact on comparability, a high level of compliance with IFRS may be of little consequence to users if this allows a low level of comparability. Similarly, a low level of compliance with IFRS may not be seen as an important issue for users when comparability is high.

Consider the first hypothesis of Nobes (2006: 237) 'International differences in practice exist among IFRS companies due to differences in the version of IFRS being used.' First, this hypothesis demands evidence that differences in practice occur. The T index is appropriate because it can take into account the degree of comparability between different practices (note that different practices can result in comparable, or partially comparable, accounts). If the comparability index is close to 1 then what these differences are due to becomes an immaterial question. Comparability indices should become a standard addition to empirical research on comparability in the same way that other summary statistics such as means and correlations are used to summarise results within a study and to enable comparisons across studies. Index values can, however, be inaccurate, especially if based on small sample sizes. The results in this paper will therefore play an important role in this research agenda by enabling standard errors and confidence intervals to accompany comparability indices.

Second, this paper enables a statistical assessment of whether the level of comparability for one topic is significantly lower than the comparability for another topic. Research effort should concentrate on topics where comparability is shown to be lower. Third, differences in comparability between different groups of countries, such as the comparison between fairness and legalistic countries suggested by Herrmann and Thomas (1995), can test theories concerning the impact of cultural and historical differences between countries.

The suggestion by Nobes (2006) that comparability may increase under IFRS also demands empirical testing. Recent evidence such as Cairns et al. (2009) suggests comparability can also significantly decrease with the adoption of IFRS. Reasons for such changes are important research questions to pursue, but investigating and proposing reasons for changes that can be attributed to random sampling variation is not a productive use of research effort. Hence the results presented in this paper will also be important for research into changes in comparability over time.

Furthermore, this paper has added insights into the characteristics of different indices within the T index framework. For example, it has been shown that the sample T index is unbiased in many cases and has negligible bias in most other practical cases. It has also illustrated how the accuracy of the T index calculated from a sample depends on the options selected for the T index and the sample sizes for each country. For example, if companies are weighted equally (option 1a) then a small sample from one country will have negligible impact on the size of the standard error compared to when countries are weighted equally (option 1b).

Finally, we have illustrated how researchers can, during early development, abandon or modify research that has little chance of achieving its aims. Previously, researchers did not have access to such valuable tools for comparability studies. Researchers can now examine whether their proposed data is likely to answer their proposed research questions prior to data collection, choose research questions and sample sizes that are realistic in advance, and report their findings using the usual statistical techniques. These include p-values to quantify evidence against hypotheses and confidence intervals and standard errors to quantify the precision of estimated levels of comparability.

Appendix

This Appendix contains four sections. The first describes the formulae to calculate the standard error of the T index. The second and third derive the formulae for the bias and variance, and hence standard error, of the T index. The last section derives the formulae for the bias and variance of the simple H, I and C special cases of the T index.

Throughout this Appendix the following notation is adopted.

N is the number of countries.
M is the number of accounting methods.
i, j, I and J are dummy indices for countries, taking integer values from 1 to N.
k, l, K and L are dummy indices for methods, taking integer values from 1 to M.
α_{kl} is the coefficient of comparability between accounting methods k and l.
β_{ij} is the weighting for the comparison between companies in countries i and j.
n_i is the number of companies sampled from country i.
X_{ki} is the number of sampled companies from country i that use method k.
p_{ki} is the proportion of sampled companies from country i that use method k.
π_{ki} is the population proportion of all companies from country i that use method k.
$E(A)$ is the expectation of A.
$Var(A)$ is the variance of A.
$Cov(A, B)$ is the covariance between A and B.

Formulae for the standard error of the T index

In this section we provide expressions for the calculation of the standard error of the T index. The standard error for the T index, σ_T, is the square-root of its variance, σ_T^2. Here we present formulae for the variance in three steps: first, as an expression involving covariances; second, expressions for these covariances in terms of expectations; and third, expressions for these expectations.

First, the variance of the T index can be expressed as

$$\sigma_T^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{M}\sum_{l=1}^{M}\sum_{I=1}^{N}\sum_{J=1}^{N}\sum_{K=1}^{M}\sum_{L=1}^{M}\alpha_{kl}\beta_{ij}\alpha_{KL}\beta_{IJ}Cov(p_{ki}p_{lj}, p_{KI}p_{LJ})$$

where Cov denotes the covariance function, i, j, I and J are dummy indices for countries, k, l, K and L are dummy indices for methods and hence, for example, p_{KI} is the proportion of companies in country I using method K. This equation has twice as many summations as are used in the definition of the T index because $Var(\sum_i X_i) = \sum_i \sum_I Cov(X_i, X_I)$ and hence an extra set of dummy indices is required. Although there is potentially a large number of terms in the above summation for σ_T^2 (N^4M^4 or 2,560,000 terms for the example in Table 1 with $N = 8$ countries and $M = 4$ methods) these terms constitute a few special cases and many of these covariances equal zero.

We now provide expressions for each of the $Cov(p_{ki}p_{lj}, p_{KI}p_{LJ})$ in the above summation in terms of expectations, denoted E, of products of up to four of these proportions. In doing so, we consider five cases depending respectively on whether 0, 1, 2, 3 or 4 of the equalities $i = I$, $i = J$, $j = I$ or $j = J$ hold.

(Case C0) There are four different country indices, so i, j, I and J are all unequal to each other.

$$Cov(p_{ki}p_{lj}, p_{KI}p_{LJ}) = 0.$$

(Case C1) There is exactly one pair of country indices that are equal to each other, so exactly one of the equalities $i = j$, $i = I$, $i = J$, $j = I$, $j = J$, or $I = J$ holds.

$$Cov(p_{ki}p_{li}, p_{KI}p_{LJ}) = 0 Cov(p_{ki}p_{lj}, p_{Ki}p_{LJ}) = E(p_{ki}p_{Ki})E(p_{lj})E(p_{LJ}) - E(p_{ki})E(p_{lj})E(p_{Ki})E(p_{LJ})$$

$$Cov(p_{ki}p_{lj}, p_{KI}p_{Li}) = E(p_{ki}p_{Li})E(p_{lj})E(p_{KI}) - E(p_{ki})E(p_{lj})E(p_{KI})E(p_{Li})$$

$$Cov(p_{ki}p_{lj}, p_{Kj}p_{LJ}) = E(p_{ki})E(p_{lj}p_{Kj})E(p_{LJ}) - E(p_{ki})E(p_{lj})E(p_{Kj})E(p_{LJ})$$

Appendix *(continued)*

$$Cov(p_{ki}p_{lj}, p_{KI}p_{Lj}) = E(p_{ki})E(p_{lj}p_{Lj})E(p_{KI}) - E(p_{ki})E(p_{lj})E(p_{KI})E(p_{Lj})$$

$$Cov(p_{ki}p_{lj}, p_{KI}p_{LI}) = 0$$

(Case C2) There are two pairs of equal country indices, so ($i = j$ and $I = J$), ($i = I$ and $j = J$) or ($i = J$ and $j = I$).

$$Cov(p_{ki}p_{li}, p_{KI}p_{LI}) = 0$$

$$Cov(p_{ki}p_{lj}, p_{Ki}p_{Lj}) = E(p_{ki}p_{Ki})E(p_{lj}p_{Lj}) - E(p_{ki})E(p_{lj})E(p_{Ki})E(p_{Lj})$$

$$Cov(p_{ki}p_{lj}, p_{Kj}p_{Li}) = E(p_{ki}p_{Li})E(p_{lj}p_{Kj}) - E(p_{ki})E(p_{lj})E(p_{Kj})E(p_{Li})$$

(Case C3) Three of the four country indices are equal to each other but the fourth country index is different, so $i = j = I \neq J$, $i = j = J \neq I$, $i = I = J \neq j$ or $j = I = J \neq i$.

$$Cov(p_{ki}p_{li}, p_{Ki}p_{LJ}) = E(p_{ki}p_{li}p_{Ki})E(p_{LJ}) - E(p_{ki}p_{li})E(p_{Ki})E(p_{LJ})$$

$$Cov(p_{ki}p_{li}, p_{KI}p_{Li}) = E(p_{ki}p_{li}p_{Li})E(p_{KI}) - E(p_{ki}p_{li})E(p_{KI})E(p_{Li})$$

$$Cov(p_{ki}p_{lj}, p_{Ki}p_{Li}) = E(p_{ki}p_{Ki}p_{Li})E(p_{lj}) - E(p_{ki})E(p_{lj})E(p_{Ki}p_{Li})$$

$$Cov(p_{ki}p_{lj}, p_{Kj}p_{Lj}) = E(p_{ki})E(p_{lj}p_{Kj}p_{Lj}) - E(p_{ki})E(p_{lj})E(p_{Kj}p_{Lj})$$

(Case C4) All four country indices are equal to each other, so $i = j = I = J$.

$$Cov(p_{ki}p_{li}, p_{Ki}p_{Li}) = E(p_{ki}p_{li}p_{Ki}p_{Li}) - E(p_{ki}p_{li})E(p_{Ki}p_{Li})$$

Finally, we provide expressions for the above expectations depending on whether we evaluate the expectation of one proportion (case E1), or the product of two (case E2), three (case E3) or four (case E4) proportions. It suffices to provide expressions for the above expectations for country i only because each expectation involves proportions for only one country. In practice, the standard error of the T index is usually estimated by substituting sample proportions p_{ki} for the population proportions π_{ki} in the expressions below.[4]

(Case E1) $E(p_{ki}) = \pi_{ki}$.

(Case E2) $E(p_{ki}p_{li})$ has two cases depending on whether the accounting methods k and l are equal.
If $k = l$, $E(p_{ki}^2) = (n_i - 1)\pi_{ki}^2/n_i + \pi_{ki}/n_i$.
If $k \neq l$, $E(p_{ki}p_{li}) = (n_i - 1)\pi_{ki}\pi_{li}/n_i$.

(Case E3) $E(p_{ki}p_{li}p_{Ki})$ has three cases depending on whether k, l and K are all equal, two are equal or none are equal to each other. Without loss of generality we consider the cases $k = l = K$, $k = l \neq K$ and k, l and K are all unequal. Note that this covers all cases since the order of the proportions is irrelevant (for example, $p_{ki}p_{li}p_{Ki} = p_{li}p_{Ki}p_{ki}$).
If $k = l = K$, $E(p_{ki}^3) = (n_i - 1)(n_i - 2)\pi_{ki}^3/n_i^2 + 3(n_i - 1)\pi_{ki}^2/n_i^2 + \pi_{ki}/n_i^2$.
If $k = l \neq K$, $E(p_{ki}^2 p_{Ki}) = (n_i - 1)(n_i - 2)\pi_{ki}^2\pi_{Ki}/n_i^2 + (n_i - 1)\pi_{ki}\pi_{Ki}/n_i^2$.
If k, l and K are all unequal, $E(p_{ki}p_{li}p_{Ki}) = (n_i - 1)(n_i - 2)\pi_{ki}\pi_{li}\pi_{Ki}/n_i^2$.

(Case E4) $E(p_{ki}p_{li}p_{Ki}p_{Li})$ has five cases depending on whether k, l, K and L are all equal, three are equal, two are equal and the other two equal each other, two are equal and the other two are different to each other, or all four are unequal. Due to symmetry (the order of the proportions is irrelevant), the following five cases cover all possibilities.

[4] Recall the important distinction between the proportion in the sample and the proportion in the population. In practice, we only know values in the sample. The formula for the standard error for a sample proportion p is $\sqrt{\pi(1-\pi)/n}$ and this is typically approximated with $\sqrt{p(1-p)/n}$. This approximation is excellent unless sample sizes are very small. We use the same approximation here by replacing π_{ki} with p_{ki} when applying these expressions for the standard error of T to sample data.

Appendix *(continued)*

If $k = l = K = L$,

$E(p_{ki}^4) = (n_i - 1)(n_i - 2)(n_i - 3)\pi_{ki}^4/n_i^3 + 6(n_i - 1)\ (n_i - 2)\pi_{ki}^3/n_i^3 + 7(n_i - 1)\pi_{ki}^2/n_i^3 + \pi_{ki}/n_i^3$.

If $k = l = K \neq L$,

$E(p_{ki}^3 p_{Li}) = (n_i - 1)(n_i - 2)(n_i - 3)\pi_{ki}^3\pi_{Li}/n_i^3 + 3(n_i - 1)(n_i - 2)\pi_{ki}^2\pi_{Li}/n_i^3 + (n_i - 1)\pi_{ki}\pi_{Li}/n_i^3$

If $k = l$ and $K = L$ with $k \neq K$,

$$E(p_{ki}^2 p_{Ki}^2) = (n_i - 1)(n_i - 2)(n_i - 3)\pi_{ki}^2\pi_{Ki}^2/n_i^3 + (n_i - 1)(n_i - 2)\pi_{ki}^2\pi_{Ki}/n_i^3 + (n_i - 1)(n_i - 2)\pi_{ki}\pi_{Ki}^2/n_i^3 + (n_i - 1)\pi_{ki}\pi_{Ki}/n_i^3.$$

If $k = l$ with $k \neq K$, $k \neq L$ and $K \neq L$,

$E(p_{ki}^2 p_{Ki} p_{Li}) = (n_i - 1)(n_i - 2)(n_i - 3)\pi_{ki}^2\pi_{Ki}\pi_{Li}/n_i^3 + (n_i - 1)(n_i - 2)\pi_{ki}\pi_{Ki}\pi_{Li}/n_i^3$.

If k, l, K and L are all unequal,

$E(p_{ki}p_{li}p_{Ki}p_{Li}) = (n_i - 1)(n_i - 2)(n_i - 3)\pi_{ki}\pi_{li}\pi_{Ki}\pi_{Li}/n_i^3$.

Derivation of the formulae for the bias of the T index

Since the bias is defined as the expected value of the sample T index minus the value of the corresponding population T index we begin by calculating this expectation. The expected value of the T index is given by

$$E(T) = E\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{M}\sum_{l=1}^{M}\alpha_{kl}\beta_{ij}p_{ki}p_{lj}\right) = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{M}\sum_{l=1}^{M}\alpha_{kl}\beta_{ij}E(p_{ki}p_{lj}) \tag{5}$$

In the appendices we repeatedly use Equation 10 of Johnson and Kotz (1969: 284) for the expectation of functions of multinomial random variables,

$E\left(X_1^{(r_1)}X_2^{(r_2)}\ldots X_R^{(r_R)}\right) = N^{(\sum_{a=1}^{R} r_a)}\pi_1^{r_1}\pi_2^{r_2}\ldots\pi_R^{r_R}$ where $A^{(r)} = A(A-1)(A-2)\ldots(A-r+1)$.

For example, $E\left(X_1^{(2)}X_2^1\right) = N(N-1)(N-2)\pi_1^2\pi_2$ and

$E\left(X_1^{(2)}X_2^1\right) = E(X_1(X_1-1)X_2) = E(X_1^2X_2) - E(X_1X_2)$.

Recall here that by definition $p_{ki} = X_{ki}/n_i$ where p_{ki} is the proportion of companies using method k in the sample from country i, X_{ki} is the sampled number of companies in country i using method k, and n_i is the sample size for country i. We therefore calculate $E(X_{ki}X_{lj})$ as the formula for $E(p_{ki}p_{lj})$ will then follow upon division by n_in_j.

There are three cases to consider for $E(p_{ki}p_{lj})$ in Equation (5) depending on whether i and j are equal (same country), and, if they are equal, depending on whether k and l are equal (same accounting method). Here we repeatedly apply Equation 10 of Johnson and Kotz (1969: 284) and the fact that $E(X_{ki}X_{lj}) = E(X_{ki})E(X_{lj})$ when $i \neq j$ because sampling is assumed independent in different countries and the expectation of a product equals the product of the expectations for two independent variables.

For $i = j$:

If $k = l$, $E(X_{ki}^2) = E(X_{ki}(X_{ki}-1)) + E(X_{ki}) = E(X_{ki}^{(2)}) + E(X_{ki}) = n_i(n_i-1)\pi_{ki}^2 + n_i\pi_{ki}$.

If $k \neq l$, $E(X_{ki}X_{li}) = n_i(n_i-1)\pi_{ki}\pi_{li}$.

For $i \neq j$: $E(X_{ki}X_{lj}) = E(X_{ki})E(X_{lj}) = n_i\pi_{ki}n_j\pi_{lj}$.

Since $p_{ik} = X_{ik}/n_i$ it follows that $E(p_{ki}p_{lj}) = E(X_{ki}X_{lj})/(n_in_j)$. Hence:

If $i = j$ and $k = l$, $E(p_{ki}^2) = (n_i-1)\pi_{ki}^2/n_i + \pi_{ki}/n_i = \pi_{ki}^2 - (\pi_{ki}^2 - \pi_{ki})/n_i$.

If $i = j$ and $k \neq l$, $E(p_{ki}p_{li}) = (n_i-1)\pi_{ki}\pi_{li}/n_i = \pi_{ki}\pi_{li} - \pi_{ki}\pi_{li}/n_i$.

For $i \neq j$: $E(p_{ki}p_{lj}) = \pi_{ki}\pi_{lj}$.

Appendix *(continued)*

From these three equations we see that $E(p_{ki}p_{lj})$ always has a term equal to $\pi_{ki}\pi_{lj}$ but has additional terms when $i = j$. Hence substituting the values for these expectations into Equation (5) yields

$$E(T) = \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{M}\sum_{l=1}^{M}\alpha_{kl}\beta_{ij}E(p_{ki}p_{lj})$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{M}\sum_{l=1}^{M}\alpha_{kl}\beta_{ij}\pi_{ki}\pi_{lj} - \sum_{i=1}^{N}\sum_{k=1}^{M}\sum_{1=1}^{M}\alpha_{kl}\beta_{ii}\pi_{ki}\pi_{li}/n_i + \sum_{i=1}^{N}\sum_{k=1}^{M}\alpha_{kk}\beta_{ii}\pi_{ki}/n_i$$

Since the first term in this expression equals the population index T_p, it follows that the bias of the T index is given, as required, by

$$\sum_{i=1}^{N}\sum_{k=1}^{M}\alpha_{kk}\beta_{ii}\pi_{ki}/n_i - \sum_{i=1}^{N}\sum_{k=1}^{M}\sum_{1=1}^{M}\alpha_{kl}\beta_{ii}\pi_{ki}\pi_{li}/n_i.$$

Derivation of the formulae for the standard error of the T index

Here we derive the formulae for the variance of the T index, σ_T^2, provided in an earlier section of this Appendix. This derivation involves three steps corresponding to the three steps in the presentation of the formulae: first, verifying the formula for σ_T^2 in terms of covariances; second, deriving the formulae for these covariances in terms of expectations (cases C0 to C4); and thirdly, deriving the expressions for these expectations (cases E1 to E4).

First, note that:

$$\sigma_T^2 = Var(T) = Var\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{M}\sum_{l=1}^{M}\alpha_{kl}\beta_{ij}p_{ki}p_{lj}\right)$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{M}\sum_{l=1}^{M}\sum_{I=1}^{N}\sum_{J=1}^{N}\sum_{K=1}^{M}\sum_{L=1}^{M}\alpha_{kl}\beta_{ij}\alpha_{KL}\beta_{IJ}Cov(p_{ki}p_{lj}, p_{KI}p_{LJ})$$

where Var and Cov denote variance and covariance functions respectively and we have repeatedly applied the formula $Var\left(\sum_{i=1}^{N} X_i\right) = \sum_{i=1}^{N}\sum_{I=1}^{N} Cov(X_i, X_I)$.

Therefore it suffices to derive the expressions for $Cov(p_{ki}p_{lj}, p_{KI}p_{LJ})$ for all possible values of i, j, I and J between 1 and N and all possible values of k, l, K and L between 1 and M. Cases C0 to C4 consider the possible values for i, j, I and J and then cases E1 to E4 consider possible cases for k, l, K and L.

Second, we derive the expressions for each of the covariances in cases C0 to C4. Each of these five cases results from whether 0, 1, 2, 3 or 4 of the equalities $i = I$, $i = J$, $j = I$ or $j = J$ hold. Below we repeatedly use the definition of covariance, $Cov(A, B) = E(AB) - E(A)E(B)$, and two fundamental rules that apply when A and B are independent: $Cov(A, B) = 0$ and $E(AB) = E(A)E(B)$. These rules apply when the terms A and B do not share information from the same country(s) because we assume that the samples of companies from the different countries are selected independently.

In particular $Cov(p_{ki}p_{lj}, p_{KI}p_{LJ}) = E(p_{ki}p_{lj}p_{KI}p_{LJ}) - E(p_{ki}p_{lj})E(p_{KI}p_{LJ})$. This expression can be simplified further when the country indices i, j, I and J are not all equal because we assume independent sampling of companies from different countries. For example, $E(p_{ki}p_{lj})$ equals $E(p_{ki})E(p_{lj})$ when $i \neq j$ and $E(p_{ki}p_{li}p_{Ki}p_{LI})$ equals $E(p_{ki}p_{li}p_{Ki})E(p_{LI})$ when $i \neq I$.

Appendix *(continued)*

(Case C0) There are four different country indices, so i, j, I and J are all unequal to each other.

If countries i and j are each different to countries I and J then the two terms $p_{ki}p_{lj}$ and $p_{KI}p_{LJ}$ are independent because we assume different countries are sampled independently. Hence $Cov(p_{ki}p_{lj}, p_{KI}p_{LJ}) = 0$.

(Case C1) There is exactly one pair of country indices that are equal to each other, so exactly one of the equalities $i = j$, $i = I$, $i = J$, $j = I$, $j = J$, or $I = J$ holds.

First, for the possibilities where either $i = j$ or $I = J$ the terms $p_{ki}p_{lj}$ and $p_{KI}p_{LJ}$ do not share information from the same country and hence are independent. Therefore

$Cov(p_{ki}p_{li}, p_{KI}p_{LJ})$ and $Cov(p_{ki}p_{lj}, p_{KI}p_{LI})$ both equal 0.

We consider the case where $i = I$. The other possibilities are derived in a similar way.

$$Cov(p_{ki}p_{lj}, p_{Ki}p_{LJ}) = E(p_{ki}p_{lj}p_{Ki}p_{LJ}) - E(p_{ki}p_{lj})E(p_{Ki}p_{LJ})$$

$$= E(p_{ki}p_{Ki})E(p_{lj})E(p_{LJ}) - E(p_{ki})E(p_{lj})E(p_{Ki})E(p_{LJ})$$

(Case C2) There are two pairs of equal country indices, so ($i = j$ and $I = J$), ($i = I$ and $j = J$) or ($i = J$ and $j = I$).

When $i = j$ and $I = J$ we have $Cov(p_{ki}p_{li}, p_{KI}p_{LI}) = 0$ because in this case the two terms $p_{ki}p_{lj}$ and $p_{KI}p_{LJ}$ are independent (if the country $i = j$ and the country $I = J$ were the same then we would have case C4 considered below).

We consider the case where $i = I$ and $j = J$. The case where $i = J$ and $j = I$ is derived in a similar way.

$$Cov(p_{ki}p_{lj}, p_{Ki}p_{Lj}) = E(p_{ki}p_{lj}p_{Ki}p_{Lj}) - E(p_{ki}p_{lj})E(p_{Ki}p_{Lj})$$

$$= E(p_{ki}p_{Ki})E(p_{lj}p_{Lj}) - E(p_{ki})E(p_{lj})E(p_{Ki})E(p_{Lj})$$

(Case C3) Three of the four country indices are equal to each other but the fourth country index is different, so $i = j = I \neq J$, $i = j = J \neq I$, $i = I = J \neq j$ or $j = I = J \neq i$.

We consider the case where $i = j = I \neq J$. The other possibilities are derived in a similar way.

$$Cov(p_{ki}p_{li}, p_{Ki}p_{LJ}) = E(p_{ki}p_{li}p_{Ki}p_{LJ}) - E(p_{ki}p_{li})E(p_{Ki}p_{LJ})$$

$$= E(p_{ki}p_{li}p_{Ki})E(p_{LJ}) - E(p_{ki}p_{li})E(p_{Ki})E(p_{LJ})$$

(Case C4) All four country indices are equal to each other, so $i = j = I = J$.

$$Cov(p_{ki}p_{li}, p_{Ki}p_{Li}) = E(p_{ki}p_{li}p_{Ki}p_{Li}) - E(p_{ki}p_{li})E(p_{Ki}p_{Li})$$

Third, we derive the expressions for the expectations specified in cases E1, E2, E3 and E4. These are the expressions for expectations depending on whether 1, 2 3 or 4 proportions appear in the expectation. Since $p_{ki} = X_{ki}/n_i$ and Equation 10 of Johnson and Kotz (1969: 284) for the expectation of functions of multinomial random variables is defined in terms of the number of companies X_{ki} we first determine the expectation of one, or a product of up to four, of the X_{ki}. The required results for the proportions p_{ki} are then obtained in each case by dividing by the appropriate n_i since, for example, $E(p_{ki}) = E(X_{ki}/n_i) = E(X_{ki})/n_i$.

(Case E1) $E(X_{ki}) = n_i\pi_{ki}$

Dividing by n_i yields, as required, $E(p_{ki}) = \pi_{ki}$.

(Case E2) $E(p_{ki}p_{li})$ has two cases depending on whether the accounting methods k and l are equal.

If $k = l$, $E(X_{ki}^2) = E(X_{ki}(X_{ki} - 1)) + E(X_{ki}) = E(X_{ki}^{(2)}) + E(X_{ki}) = n_i(n_i - 1)\pi_{ki}^2 + n_i\pi_{ki}$

If $k \neq l$, $E(X_{ki}X_{li}) = n_i(n_i - 1)\pi_{ki}\pi_{li}$

Appendix *(continued)*

Dividing each of these expressions by n_i^2 yields, as required,

If $k = l$, $E(p_{ki}^2) = (n_i - 1)\pi_{ki}^2/n_i + \pi_{ki}/n_i$.

If $k \neq l$, $E(p_{ki}p_{li}) = (n_i - 1)\pi_{ki}\pi_{li}/n_i$

(Case E3) $E(p_{ki}p_{li}p_{Ki})$ has three cases depending on whether k, l and K are all equal, two are equal or none are equal to each other.

If $k = l = K$,

$$E(X_{ki}^3) = E(X_{ki}(X_{ki} - 1)(X_{ki} - 2)) + 3E(X_{ki}(X_{ki} - 1)) + E(X_{ki})$$
$$= E(X_{ki}^{(3)}) + 3E(X_{ki}^{(2)}) + E(X_{ki}) = n_i(n_i - 1)(n_i - 2)\pi_{ki}^3 + 3n_i(n_i - 1)\pi_{ki}^2 + n_i\pi_{ki}$$

If $k = l \neq K$

$$E(X_{ki}^2X_{Ki}) = E(X_{ki}(X_{ki} - 1)X_{Ki}) + E(X_{ki}X_{Ki}) = E(X_{ki}^{(2)}X_{Ki}) + E(X_{ki}X_{Ki})$$
$$= n_i(n_i - 1)(n_i - 2)\pi_{ki}^2\pi_{Ki} + n_i(n_i - 1)\pi_{ki}\pi_{Ki}$$

If k, l and K are all unequal

$$E(X_{ki}X_{li}X_{Ki}) = n_i(n_i - 1)(n_i - 2)\pi_{ki}\pi_{li}\pi_{Ki}$$

Dividing each of these expressions by n_i^3 yields the required results.

(Case E4) $E(p_{ki}p_{li}p_{Ki}p_{Li})$ has five cases depending on whether k, l, K and L are all equal, three are equal, two are equal and the other two equal each other, two are equal and the other two are different to each other, or all four are unequal.

If $k = l = K = L$,

$$E(X_{ki}^4) = E(X_{ki}(X_{ki} - 1)(X_{ki} - 2)(X_{ki} - 3)) + 6E(X_{ki}(X_{ki} - 1)\ (X_{ki} - 2)) + 7E(X_{ki}(X_{ki} - 1))$$
$$+E(X_{ki}) = E(X_{ki}^{(3)}) + 3E(X_{ki}^{(2)}) + E(X_{ki})$$
$$= n_i(n_i - 1)(n_i - 2)(n_i - 3)\pi_{ki}^4 + 6n_i(n_i - 1)(n_i - 2)\pi_{ki}^3 + 7n_i(n_i - 1)\pi_{ki}^2 + n_i\pi_{ki}$$

If $k = l = K \neq L$,

$$E(X_{ki}^3X_{Li}) = E(X_{ki}(X_{ki} - 1)(X_{ki} - 2)X_{Li}) + 3E(X_{ki}(X_{ki} - 1)X_{Li}) + E(X_{ki}X_{Li})$$
$$= E(X_{ki}^{(3)}X_{Li}) + 3E(X_{ki}^{(2)}X_{Li}) + E(X_{ki}X_{Li})$$
$$= n_i(n_i - 1)(n_i - 2)(n_i - 3)\pi_{ki}^3\pi_{Li} + 3n_i(n_i - 1)(n_i - 2)\pi_{ki}^2\pi_{Li} + n_i(n_i - 1)\pi_{ki}\pi_{Li}$$

If $k = l$ and $K = L$ with $k \neq K$,

$$E(X_{ki}^2X_{Ki}^2) = E(X_{ki}(X_{ki} - 1)X_{Ki}(X_{Ki} - 1)) + E(X_{ki}(X_{ki} - 1)X_{Ki}) + E(X_{ki}X_{Ki}(X_{Ki} - 1))$$
$$-E(X_{ki}X_{Ki}) = E(X_{ki}^{(2)}X_{Ki}^{(2)}) + E(X_{ki}^{(2)}X_{Ki}) + E(X_{ki}X_{Ki}^{(2)}) + E(X_{ki}X_{Ki})$$
$$= n_i(n_i - 1)(n_i - 2)(n_i - 3)\pi_{ki}^2\pi_{Ki}^2 + n_i(n_i - 1)(n_i - 2)\pi_{ki}^2\pi_{Ki}$$
$$+n_i(n_i - 1)(n_i - 2)\pi_{ki}\pi_{Ki}^2 + n_i(n_i - 1)\pi_{ki}\pi_{Ki}$$

If $k = l$ with $k \neq K$, $k \neq L$ and $K \neq L$,

$$E(X_{ki}^2X_{Ki}X_{Li}) = E(X_{ki}(X_{ki} - 1)X_{Ki}X_{Li}) + E(X_{ki}X_{Ki}X_{Li}) = E(X_{ki}^{(2)}X_{Ki}X_{Li}) + E(X_{ki}X_{Ki}X_{Li})$$
$$= n_i(n_i - 1)(n_i - 2)(n_i - 3)\pi_{ki}^2\pi_{Ki}\pi_{Li} + n_i(n_i - 1)(n_i - 2)\pi_{ki}\pi_{Ki}\pi_{Li}$$

If k, l, K and L are all unequal,

$$E(X_{ki}X_{li}X_{Ki}X_{Li}) = n_i(n_i - 1)(n_i - 2)(n_i - 3)\pi_{ki}\pi_{li}\pi_{Ki}\pi_{Li}$$

Dividing these expressions by n_i^4 yields the required results.

Appendix *(continued)*

Derivation of bias and variance of the H, I and C special cases of the T index

Here we derive the formulae for the bias and variance of the simple indices in Table 3. The bias and variance for the H index (options 1a2a3a4a) when there is one country are derived in Taplin (2003).

For the within-country index (options 1a2b3a4a) $C_w = \sum_i \beta_{ii} H_i$, where H_i is the H index applied to country i. Hence the bias is a weighted sum of the bias for the H_i and the variance $Var(C_w) = \sum_i \beta_{ii}^2 Var(H_i)$, as required, because the samples from the different countries are taken independently.

For the two-country I index (options 1a2c3a4a) for countries i and j we have

$$I_{ij} = \sum_k p_{ki}p_{kj} \text{ so } E(I_{ij}) = \sum_k \pi_{ki}\pi_{kj} \text{ and } E(I_{ij})^2 = \sum_k \sum_l \pi_{ki}\pi_{kj}\pi_{li}\pi_{lj}.$$

$$E(I_{ij}^2) = E(\sum_k \sum_l p_{ki}p_{kj}p_{li}p_{lj}) = \sum_k \sum_l E(p_{ki}p_{kj}p_{li}p_{lj}) = \sum_k \sum_l E(p_{ki}p_{li})E(p_{kj}p_{lj}).$$

Hence

$$Var(I_{ij}) = E(I_{ij}^2) - E(I_{ij})^2 = \sum_k \sum_l \left(E(p_{ki}p_{li})E(p_{kj}p_{lj}) - \pi_{ki}\pi_{kj}\pi_{li}\pi_{lj}\right) = \sum_k \sum_l \theta_{kl(ij)}$$

where $\theta_{kl(ij)}$ is derived for $k \neq l$ from Equation (4.2) of Johnson and Kotz (1969: 51) to be:

$$E(p_{ki}p_{li})E(p_{kj}p_{lj}) - \pi_{ki}\pi_{kj}\pi_{li}\pi_{lj} = ((n_i - 1)\pi_{ki}\pi_{li})\big((n_j - 1)\pi_{kj}\pi_{lj}\big)/(n_i n_j) - \pi_{ki}\pi_{kj}\pi_{li}\pi_{lj}$$

$= (1 - n_i - n_j)\pi_{ki}\pi_{kj}\pi_{li}\pi_{lj}/(n_i n_j)$ and is derived for $k = l$ from Equation (4.41) of Johnson and Kotz (1969: 51) to be:

$$E(p_{ki}^2)E(p_{kj}^2) - \pi_{ki}\pi_{kj}\pi_{li}\pi_{lj} = \Big((n_i - 1)\pi_{ki}^2 + \pi_{ki}\Big)\Big((n_j - 1)\pi_{kj}^2 + \pi_{kj}\Big)/(n_i n_j) - \pi_{ki}^2\pi_{kj}^2,$$

as required.

For the between-country C index with more than two countries (options 1a2c3a4a) we have

$$C_b = \sum_i \sum_{j \neq i} \beta_{ij} \sum_k p_{ki}p_{kj},$$

$E(C_b) = \sum_i \sum_{j \neq i} \beta_{ij} \sum_k \pi_{ki}\pi_{kj}$ and

$$E(C_b)^2 = \sum_i \sum_{j \neq i} \sum_I \sum_{J \neq I} \beta_{ij}\beta_{IJ} \sum_k \sum_l \pi_{ki}\pi_{kj}\pi_{lI}\pi_{lJ}.$$

Since
$Var(C_b) = E(C_b^2) - E(C_b)^2 = \sum_i \sum_{j \neq i} \sum_k \sum_l \sum_I \sum_{J \neq I} \beta_{ij}\beta_{IJ}\Big(E(p_{ki}p_{kj}p_{lI}p_{lJ}) - \pi_{ki}\pi_{kj}\pi_{lI}\pi_{lJ}\Big)$ the desired result follows by considering several special cases depending on whether I or J equal either i or j.

First, if i, j, I, and J are all unequal to each other then by independence of sampling in different countries $\beta_{ij}\beta_{IJ}\big(E(p_{ki}p_{kj}p_{lI}p_{lJ}) - \pi_{ki}\pi_{kj}\pi_{lI}\pi_{lJ}\big) = 0$.

Second, if $i = I$ and $j = J$ (or alternatively when $i = J$ and $j = I$),

$$\beta_{ij}\beta_{IJ}\big(E(p_{ki}p_{kj}p_{lI}p_{lJ}) - \pi_{ki}\pi_{kj}\pi_{lI}\pi_{lJ}\big) = \beta_{ij}^2\big(E(p_{ki}p_{kj}p_{li}p_{lj}) - \pi_{ki}\pi_{kj}\pi_{li}\pi_{lj}\big) = \beta_{ij}^2\theta_{kl(ij)}$$

by definition of $\theta_{kl(ij)}$. This explains the term $\theta_{kl(ij)}$ in the result, with the 2 in the formula because there are two combinations of I and J yielding $\theta_{kl(ij)}$.

Appendix *(continued)*

Third, if $i = I$ but j and J differ from each other and from $i = I$, then

$$\sum_{J \neq I} \beta_{ij}\beta_{IJ}\big(E(p_{ki}p_{kj}p_{lI}p_{lJ}) - \pi_{ki}\pi_{kj}\pi_{lI}\pi_{lJ}\big) = \sum_{J \neq I} \beta_{ij}\beta_{iJ}\big(E(p_{ki}p_{li})E(p_{kj})E(p_{lJ}) - \pi_{ki}\pi_{kj}\pi_{li}\pi_{lJ}\big)$$

$$= \beta_{ij}\pi_{kj}\pi_{lJ}\sum_{J \neq I}\beta_{iJ}(E(p_{ki}p_{li}) - \pi_{ki}\pi_{li})$$

and $E(p_{ki}p_{li}) - \pi_{ki}\pi_{li}$ equals, once again using the formula from Johnson and Kotz (1969), $-\pi_{ki}\pi_{kj}\pi_{li}\pi_{lJ}/n_i$ when $k \neq l$ and equals $\pi_{ki}\pi_{kj}(1 - \pi_{li})\pi_{lJ}/n_i$ when $k = l$.

Fourth, in a similar derivation if $j = I$ but i and J differ from each other and from $j = I$, then

$$\sum_{J \neq I} \beta_{ij}\beta_{IJ}\big(E(p_{ki}p_{kj}p_{lI}p_{lJ}) - \pi_{ki}\pi_{kj}\pi_{lI}\pi_{lJ}\big) = \sum_{J \neq I} \beta_{ij}\beta_{jJ}\big(E(p_{kj}p_{lj})E(p_{ki})E(p_{lJ}) - \pi_{ki}\pi_{kj}\pi_{lj}\pi_{lJ}\big)$$

$$= \beta_{ij}\pi_{ki}\pi_{lJ}\sum_{J \neq I}\beta_{jJ}\big(E(p_{kj}p_{lj}) - \pi_{kj}\pi_{lj}\big)$$

and $E(p_{kj}p_{lj}) - \pi_{kj}\pi_{lj}$ equals, once again using the formula from Johnson and Kotz (1969), $-\pi_{ki}\pi_{kj}\pi_{lj}\pi_{lJ}/n_j$ when $k \neq l$ and equals $\pi_{ki}\pi_{kj}(1 - \pi_{lj})\pi_{lJ}/n_j$ when $k = l$.

Adding the third and fourth cases above we obtain

$\phi_{kl(ij)} = \pi_{ki}\pi_{kj}\sum_{J \neq i,j}(\pi_{li}\pi_{lJ}\beta_{iJ}/n_i + \pi_{lj}\pi_{lJ}\beta_{jJ}/n_j)$ when $k \neq l$ and

$\phi_{kk(ij)} = \pi_{ki}\pi_{kj}\sum_{J \neq i,j}(\pi_{kJ}(\pi_{ki} - 1)\beta_{iJ}/n_i + \pi_{kJ}(\pi_{kj} - 1)\beta_{jJ}/n_j)$.

There are only two further possible combinations for i, j, I, and J. These are when $i = J$ but j and I differ from each other and from $i = J$, which gives the same results as the third case above (except the dummy index I replaces the dummy index J), and when $j = J$ but i and I differ from each other and from $j = J$, which gives the same result as the fourth case. Thus each $\phi_{kl(ij)}$ is required twice, as was the case for the $\theta_{kl(ij)}$, and hence the 2 in the formula for the variance of C_b.

References

Ali, M.J. (2005). 'A synthesis of empirical research on international accounting harmonization and compliance with international financial accounting standards'. *Journal of Accounting Literature*, 24: 1–52.

Archer, S., Delvaille, P and Mcleay, S. (1995). 'The measurement of harmonisation and the comparability of financial statement items: within-country and between-country effects'. *Accounting and Business Research*, 25 (Spring): 67–80.

Astami, E.W., Tower, G., Rusmin and Taplin, R. (2006). 'An exploration of the attributes of international harmonisation indices using the T-Index'. *The Journal of Contemporary Issues in Business and Government*, 12(2): 47–66.

Barth, M.E., Clinch, G. and Shibano, T. (1999). 'International accounting harmonization and global equity market'. *Journal of Accounting & Economics*, 26: 201–235.

Cairns, D, Massoudi, D.R., Taplin, R. and Tarca, A. 'IFRS fair value measurement and accounting policy choice in the United Kingdom and Australia'. (6 February 2009). AAA 2009 Mid-Year International Accounting Section (IAS) Meeting. Available at SSRN: http://ssrn.com/abstract=1274024.

Cole, V., Branson, J. and Breesch, D. (2008). 'Finding a method to measure the comparability of the consolidated financial statements of the European listed companies from the viewpoint of the users'. Working paper presented at the 31st European Accounting Association (EAA) Annual Congress, April 23–25, 2008, Rotterdam, Netherlands.

Herrmann, D. and Thomas, W. (1995). 'Harmonisation of accounting measurement practices in the European Community'. *Accounting and Business Research*, 25 (Autumn): 253–265.

Jaffar, A. and McLeay, S. (2007). 'Country effects and sector effects on the harmonization of accounting policy choice'. *ABACUS*, 43(2): 156–189.

Johnson, N.L. and Kotz, S. (1969). *Distributions in Statistics: Discrete Distributions*. Wiley.

Morris, R.D. and Parker, R.H. (1998). 'International harmony measures of accounting policy: comparative statistical properties'. *Accounting and Business Research*, 29 (Winter): 73–86.

Pierce, A. and Weetman, P. (2002). 'Measurement of de facto harmonization: implications of non-disclosure for research planning and interpretation'. *Accounting and Business Research*, 32(4): 259–273.

Taplin, R.H. (2003). 'Harmony, statistical inference with the Herfindahl H Index and C Index'. *ABACUS*, 39(2): 82–94.

Taplin, R.H. (2004). 'A unified approach to the measurement of international accounting harmony'. *Accounting and Business Research*, 34 (Spring): 57–73.

Taplin, R.H. (2006). 'A practical guide to constructing specialized indices of international accounting harmony using the harmoniser'. *Financial Reporting Regulation and Governance*. 5(1): 1–28.

Tay, J.S.W. and Parker, R.H. (1990). 'Measuring international harmonization and standardisation'. *ABACUS*, 26 (1): 71–88.

van der Tas, L.G. (1988). 'Measuring harmonisation of financial reporting practice'. *Accounting and Business Research*, 18 (Spring): 157–169.