



## Cross-country cross-survey design in international marketing research: The role of input data in multiple imputation

Journal:	<i>International Marketing Review</i>
Manuscript ID:	IMR-11-2014-0348.R1
Manuscript Type:	Original Article
Keywords:	international marketing research, multiple imputation, split questionnaire design, cross-survey, cross-country, measurement invariance

SCHOLARONE™  
Manuscripts

Review

## CROSS-COUNTRY CROSS-SURVEY DESIGN IN INTERNATIONAL MARKETING RESEARCH: THE ROLE OF INPUT DATA IN MULTIPLE IMPUTATION

### Abstract

Purpose – The present paper focuses on the case where – by design – one needs to impute cross-country cross-survey data (situation typical for example among multinational firms who are confronted with the need to carry out comparative marketing surveys with respondents located in several countries). Importantly, while some work demonstrates approaches for single-item direct measures, no prior research has examined the common situation in international marketing where the researcher needs to use multi-item scales of latent constructs. Our paper presents problem areas related to the choices international marketers have to make when doing cross-country / cross-survey research and provides guidance for future research.

Design/methodology/approach – Multi-country sample of real data is used as an example of cross-sample imputation (292 New Zealand exporters and 302 Finnish ones) the international entrepreneurial orientation data. Three variations of the input data are tested: A) imputation based on all the data available for the measurement model, B) imputation based on the set of items based on the invariance structure of the joint items shared across the two groups, and C) imputation based both on examination of the invariance structures of the joint items and the performance of the measurement model in the group where the full data was originally available.

Findings – Based on distribution comparisons imputation for New Zealand after completing the measurement model with Finnish data (Model C) gave the most promising results.

Consequently, using knowledge on between country measurement qualities may improve the imputation results, but this benefit comes with a downside since it simultaneously reduces the amount of data used for imputation. None of the imputation models leads to the same

1  
2  
3 statistical inferences about covariances between latent constructs than as the original full data,  
4 however.  
5  
6  
7

8  
9 Research limitations / Implications - The present exploratory study suggests that there are  
10 several concerns and issues that should be taken into account when planning cross-country  
11 cross-surveys. These concerns arising from current study lead us to question the  
12 appropriateness of the cross-country cross-survey approach in general although in general  
13 advantages exist. Further research is needed to find the best methods.  
14  
15  
16  
17

18  
19 Originality / value – The combination of cross-country and cross-survey approaches is novel to  
20 international marketing, and it is not known how the different procedures utilized in imputation  
21 affect the results and their validity and reliability. We demonstrate the consequences of the  
22 various imputation strategy choices taken by using a real example of a two-country sample.  
23  
24 Our exploration may have significant implications to international marketing researchers and  
25 the paper offers stimulus for further research in the area.  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## Cross-country cross-survey design in international marketing research: The role of input data in multiple imputation

### Introduction

As markets and marketing activities have become more integrated and global in scope, there is increasing interest and activity among both academics and practitioners around conducting research in multi-country settings (Douglas and Craig, 2006; He *et al.*, 2008). However, multi-country research designs should not be thought of as a simple extension of existing single-country research. Rather, there are significant practical and technical difficulties regarding cross-country research that must be dealt with. For example, in a practical sense, data collection costs are likely to be higher due to issues like questionnaire translation and back-translation procedures. From a technical standpoint, significant analytic and research design issues have emerged in the last two decades regarding the challenges of drawing meaningful conclusions from multi-country data (e.g. Steenkamp and Baumgartner, 1998; Oliveira, Cadogan, and Souchon, 2012; Franke and Richey, 2010).

Thanks to research such as that cited above, researchers should be aware of many of the key complications that researchers face in conducting multi-country research. However, underlying these problems “are more subtle issues that threaten the integrity of research” (Douglas and Craig, 2006, p.1). In the present paper, we attempt to address one subset of these more subtle issues; the use of multiple imputation methods in cross-country research, in the situation where individual samples have been taken from multiple countries. We explore and describe a number of different approaches to how this imputation can be done and discuss the consequences of the choices an international marketing researcher can make when designing his/her survey study. The starting point in our exploration is a *cross-country / cross-survey* approach. In such a case, multiple samples are taken (i.e. cross-country), but each sample only has a subset of items in common with the others (i.e. cross-survey). Cross-country research is already well established, and cross-survey research has also received attention in marketing. However, the combination of these two approaches is novel to international marketing, and although it may have the potential to help ameliorate a number of key problems

1  
2  
3 currently facing international marketing research it is not yet known how the choice of specific  
4 procedures utilized in the imputation may affect the validity and reliability of research results.  
5  
6  
7

8 While there is an array of diverse methods available to international business researchers (e.g.  
9 experiments, secondary data studies, qualitative ethnographic work, and so on) it seems that  
10 international business and marketing research commonly employs survey-based methods. For  
11 example, Chang et al. (2010) report that 40% of articles in the *Journal of International*  
12 *Business Studies* use survey methods, and our own survey of articles in the *International*  
13 *Marketing Review* finds that around 50% of articles published from 2010-2013 use survey  
14 methods<sup>1</sup>. As such, like all survey research, international marketing faces significant issues of  
15 low sample sizes, compounded by low and falling response rates, potentially leading to serious  
16 bias in research results. A key cause of low response rates is the questionnaire length, which  
17 has been shown to affect the quality of the data collected in several ways (Berdie, 1989;  
18 Herzog and Bachman, 1980; Adams and Darwin, 1982; Dillman *et al.*, 1993; Roszkowski and  
19 Bean, 1990). As well as potentially increasing nonresponse bias, low response rates are  
20 problematic as they may result in a loss of power to detect effects, due to a resulting  
21 inadequate sample size, and inaccurate effect size estimation. Small samples also limit our  
22 chances of testing larger and/or more complex models (such as models with interaction terms,  
23 or other non-linear relationships). Further, where this is important, low response rates may  
24 play havoc with the representativeness of a sample. All these problems create issues with  
25 testing, validity, reliability, international comparative work, and generalization.  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41

42 While questionnaire length is an issue for all research designs that use questionnaires, it is of  
43 special concern for researchers dealing with cross-country data. This concern arises in part  
44 from the cross-sample invariance requirements that are inherent to cross-country studies  
45 (Steenkamp and Baumgartner, 1998). In contemporary times, many journals have correctly  
46 adopted stringent policies on such issues (c.f. He *et al.*, 2008), and require extensive evidence  
47 for the invariance of measures across samples, before any claims regarding cross-sample  
48 differences or similarities can be made. However, establishing cross-cultural invariance is  
49  
50  
51  
52  
53  
54

---

55  
56  
57 <sup>1</sup> Full results of this analysis of *IMR* articles 2010-2013 are omitted due to space constraints, but are available on request  
58 from the authors.  
59  
60

1  
2  
3 often challenging (De Jong *et al.*, 2007) and it is rare for researchers to achieve total  
4 invariance of all measures. Of course, given the standard procedures for achieving full or  
5 partial invariance in multi-item measures, the process is much more likely to be successful if  
6 measures originally contain a relatively large (over 7 or 8) number of items. However, this  
7 consideration is somewhat at odds with the need to keep questionnaire length to the minimum  
8 – particularly where complex models are being examined.  
9

10  
11  
12  
13  
14  
15 Given these issues, it is clear that scholars need to find ways to increase both response rates  
16 and in turn ultimate sample sizes, in cross-national research situations. One way of doing this  
17 is to explore the use of planned missing data designs (e.g. see Enders, 2010; Graham *et al.*,  
18 2006 for introductory general reviews), like cross-survey design, sub-sampling, split  
19 questionnaire design, or data fusion. However, while such approaches look to be especially  
20 attractive for scholars conducting cross-country studies, there is little advice in the extant  
21 literature on how to deal with cross-survey / cross-country data, which (by design) contain a  
22 significant amount of missing data. Littvay (2009) has provided some general advice however,  
23 particularly concerning cross-survey sampling design issues. Further, the latter makes key  
24 recommendations concerning different missing data patterns that can be used.  
25  
26  
27  
28  
29  
30  
31  
32

33  
34 However, cross-country studies bring forth an additional dimension to cross-survey designs.  
35 Specifically, in international marketing research, it is assumed that in cross-country research,  
36 country-specific samples represent the same population and are thus drawn from the common  
37 universe (Rendall *et al.*, 2013). This assumption needs to be carefully studied prior to data  
38 collection. In addition, measurement invariance issues are of high importance when cross-  
39 country data is analyzed. Measurement invariance is a critical concept in any research context  
40 where multiple groups are sampled, and such situations are more common than many  
41 researchers may realise. For example, many surveys can be divided into distinct groups (e.g.  
42 gender, education status, job type), and the researcher must be certain that such groups are  
43 unlikely to exhibit measurement variance if they wish to consider the sample as a single pool.  
44 In the present study, we restrict ourselves to discussing the case of cross-national research,  
45 since it is within international marketing and business research that issues of measurement  
46 invariance have been most thoroughly covered. However, the key concepts and techniques we  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 present here are in principle applicable to many other research contexts, and all researchers  
4 would be advised to take these issues into account.  
5  
6  
7  
8  
9

10 Typically, measurement invariance is approached during the measurement validation phase of  
11 the analytical procedure (e.g. Steenkamp and Baumgartner, 1998), and obtaining invariant  
12 measures and valid measurement models often may require significant modifications of the  
13 original measurement model. Thus, all the items collected to represent a latent variable are not  
14 necessarily used at the end in the final model. It is at this point that practices of cross-country  
15 research in terms of measurement invariance contradict practices of multiple imputation. It is  
16 generally recommended that when imputing data, all the relevant data should be used at the  
17 imputation phase (Reiter *et al.*, 2006; Schafer and Graham, 2001). However, in cross-national  
18 research contexts, it is often difficult to know which items will be included in the final model, as  
19 the measurement model might need some modifications to ensure the invariance of measures  
20 across the groups (or countries). Unfortunately, existing research has no advice currently on  
21 how to address such issues in cross-country cross-survey designs. In other words, for  
22 example, are the outcomes of multiple imputation methods dependent on the level of  
23 invariance of the input variables, or more broadly stated: how does the input data affect the  
24 results of multiple imputation in cross-country research?  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

#### 40 **Purpose of the article**

41  
42  
43 In the present paper, we focus on the case where – by design – one needs to impute cross-  
44 country cross-survey data (a situation typical among, for example multinational firms who are  
45 confronted with the need to carry out comparative marketing surveys with respondents located  
46 in several countries), and where the researcher also wishes to use multi-item scales of latent  
47 constructs. Importantly, while some work demonstrates approaches for single-item direct  
48 measures, no prior research has examined the common situation in international marketing  
49 where the researcher needs to use multi-item scales of latent constructs. In such cases, cross-  
50 sample invariance is also a critical consideration, and no prior work is able to advise  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 researchers on how to take account of such issues while still utilizing cross-survey / cross-  
4 national designs with imputation. As such, our exploration is novel and may have significant  
5 implications to international marketing researchers. We demonstrate the consequences of the  
6 various imputation strategy choices by using a real example of a two-country sample with a  
7 measure of entrepreneurial orientation<sup>2</sup>. Although all the items were originally included in data  
8 collection for both countries and we show the results of the full confirmatory factor analyses  
9 and invariance testing results, we provide three hypothetical illustrations based on the use of  
10 different input data in multiple imputations. This would replicate the situation where only a  
11 handful of items are in common, yet items needed for theory testing are shared partially across  
12 the country samples, and enables us to describe and evaluate the different multiple imputation  
13 designs as a tool for complementing cross-country / cross-survey data that has missing data  
14 by design. Such a case looks to be an archetypal example of how a cross-country / cross-  
15 survey design could help researchers reduce questionnaire length, and thus increase  
16 response rate.  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31

### 32 **Fundamentals of missing data and multiple imputation**

33  
34  
35  
36 In general, missing data is a serious problem for researchers. That said, researchers have  
37 typically shown greater concern for data that is missing unintentionally, for example due to  
38 respondents' reluctance to provide answers to sensitive issues (e.g. Gottschall *et al.*, 2012).  
39 However, there exist cases where data can be missing intentionally, for example in situations  
40 where researchers would like to combine two independently administered data sets (like in  
41 split questionnaire designs, in sub-sampling, or data fusion), or due to other reasons  
42 irrespective from the respondents.  
43  
44  
45  
46  
47  
48  
49

50 Combining the two data sets is rather straightforward and easy, but the problem arises as we  
51 look at the merged data, as it has a lot of empty cells due to the differences in the original data  
52 sets. These missing values have two major negative effects: they have a negative effect on  
53  
54

---

55  
56  
57 <sup>2</sup> We use two countries for simplicity's sake. However, our method is not restricted to two-sample situations, and is  
58 extensible in principle to any number of samples.  
59  
60



1  
2  
3 statistical power and they may result in biased estimates (Tsiriktsis, 2005), and thus, it is  
4 recommended to 'fill in' the missing data (as deleting missing cases is not recommended or  
5 even possible here). There exist traditional techniques to fill in the holes, like different  
6 replacement procedures, but they have serious limitations (Olinsky *et al.*, 2003). Recently,  
7 more sophisticated techniques for dealing with missing values are developed, and they are  
8 known as multiple imputation methods. However, before any actions can be taken for 'filling in'  
9 missing data with plausible values, researchers need to verify the distribution of missingness  
10 for their case.  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20

### 21 *The Distribution of missingness*

22  
23 When considering imputation, the single most important question that a researcher must  
24 address is whether the pattern of missing observations is random or not. Data can be either  
25 missing at random (MAR), missing completely at random (MCAR), or not missing at random  
26 (NMAR, see Little and Rubin, 1987). These terms describe relationships between measured  
27 variables and the probability of missing data. NMAR refers to a situation, where probability of  
28 missing data is systematically related to the hypothetical values of the variables (but which are,  
29 of course missing). For example, if managing directors of poorly performing firms skip the  
30 questions about firm performance because of reluctance to report their poor performance, the  
31 pattern of missingness is NMAR. If the missing data pattern is NMAR then there is no  
32 statistical means to alleviate the problem. MAR has less stringent assumption about the  
33 reason for missing data. In MAR missingness is related to other measured variables in the  
34 analysis model, but not to the underlying values of the incomplete variable. Consider a  
35 researcher who is interested in comparing family businesses and other businesses in terms of  
36 firm performance. She defines family business as a business where family owns more than 50  
37 percent of shares, and collects this ownership data with survey and combines it with objective  
38 performance data on publicly listed companies. The majority of family owned firms are not,  
39 however, publicly listed, which means that performance data for these companies are missing.  
40 Thus, the probability of missing performance data is systematically related to values of  
41 ownership data but unrelated to hypothetical values that are missing, which qualifies as MAR.  
42 According to Olinsky *et al.* (2003), many multiple imputation methods assume that pattern of  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 missingness is MAR. In MCAR missingness is completely unsystematic and the observed data  
4 can be thought of as a random subsample of the hypothetically complete data. When pattern  
5 of missingness is MCAR regression analyses, means, nonparametric tests and moment-based  
6  
7 techniques are valid without imputing missing values (Little, 1988; Olinsky *et al.*, 2003). MCAR  
8  
9 missingness also satisfies assumptions for imputing missing values. Thus, MCAR and MAR  
10  
11 missingness patterns satisfy the conditions of using multiple imputation methods.  
12  
13  
14

15  
16 According to Rendall *et al.* (2013), cross-survey missingness is monotonic and easily satisfies  
17 the MAR assumption needed for unbiased multiple imputation. For example, if two surveys,  
18 Survey 1 with fewer variables (X1, X2 and X3) and Survey 2 with more variables (X1, X2, X3,  
19 X4, and X5) are collected separately and combined. In the matched data the probability of  
20 missing values for X4 and X5 are related to the survey from which the observation was  
21 collected (because the questions were not presented), and not to the hypothetical values that  
22 are missing. This satisfies MAR assumption, if the surveys sample from a common universe  
23 using invariant instruments. So MAR describes systematic missingness. MCAR data may also  
24 be a purposeful due to the research design (Fichman and Cummings, 2003) - like due to split  
25 questionnaire design. For example, when researcher creates two versions of the  
26 questionnaire, where Questionnaire 1 measures set of variables (X1, X2, X3, and X4) and  
27 Questionnaire 2 measures another set of variables (X1, X2, X3, and X5), and randomly  
28 assigns these questionnaires to respondents. The probabilities of missing data for X4, and X5  
29 are unrelated to any of the variables in the dataset. The missing-by-design structure of the  
30 pooled observations used in cross-survey imputations implies that missingness has a  
31 monotone rather than arbitrary pattern (Rubin, 1987).  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44

#### 45 46 *Dealing with missing data*

47 Many different missing data replacement procedures exist (c.f. e.g. Rubin (1986) and  
48 Tsiriktsis (2005) for a comparison), such as listwise deletion, maximum likelihood (ML)  
49 methods, expectation maximization (EM), and multiple imputation (MI). Multiple imputation has  
50 emerged as a flexible alternative to likelihood methods for a wide variety of missing-data  
51 problems (Schafer and Graham, 2002) (see Rubin, 1996; Schafer, 1997 for a review). We do  
52 not claim to review and compare all different missing data procedures, and refer readers to  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 dedicated sources for this (e.g. Schafer *et al.* (2002) for a general state of the art review on  
4 missing data, and Brown (1994) for comparison of five methods for modeling missing values in  
5 structural equation modeling). Rather, we focus on describing one specific approach in detail.  
6  
7 Our approach is dedicated to the specific needs of international marketers by exploiting an  
8 existing method for dealing with the missing data allowing the further use of standard analyses  
9 methods, like structural equation modeling or other multivariate analysis methods.  
10  
11  
12  
13  
14

### 15 *Multiple imputation methods for dealing with missing items*

16 Multiple imputation (MI) is a common method to deal with missing data problems (Honaker and  
17 King, 2010). MI methods have been developed primarily for imputation from complete cases to  
18 incomplete cases in the same survey ("within-survey MI") but have been applied also in a  
19 cross-survey manner (Rendall *et al.*, 2013). Multiple imputation is attractive as it works in  
20 conjunction with standard software allowing researchers to carry out analyses using SAS,  
21 LISREL, or virtually any other statistical package (Schafer and Olsen, 1998), and is thus used  
22 here as currently multivariate methods and structural equation modeling is often applied in  
23 international marketing research.  
24  
25  
26  
27  
28  
29  
30  
31  
32

33 The typical multiple imputation process consists of three steps: imputing the data, analyzing  
34 the data and pooling the results (Baraldi and Enders, 2010). During the imputing phase several  
35 copies of the data set are created, each containing different imputed values. Multiple  
36 imputation is based on replacing each missing value by a list of  $m > 1$  simulated values  
37 (Schafer and Graham, 2002), thus, creating new imputed data sets. In each of these data sets,  
38 the observed values are the same, and the imputations vary depending on the estimated  
39 uncertainty in predicting each missing value (Honaker and King, 2010). The results are then  
40 combined in specific ways for analysis purposes (Rubin, 1987; Schafer, 1997). This is followed  
41 by the analysis phase based on the pooled estimates.  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52

### 53 **Special features of cross-country cross-survey designs**

54  
55  
56  
57  
58  
59  
60

1  
2  
3 However, things are much more complicated with cross-survey data consisting of multi-item  
4 latent variables collected from different samples (or groups, like countries) and analyzed using  
5 structural equation modeling. For example, it has been proposed that multiple imputed data  
6 sets can be analyzed in SEM programs using the multiple group approach (Verleye, 1997).  
7 Unfortunately, the multi-group approach for MI often becomes impossible in the international  
8 marketing context, because of the multi-group approach also required by cross-country data.  
9 In addition, in cross-country studies several aspects of invariance need to be dealt with  
10 (Malhotra *et al.*, 1996), like sampling invariance before the data collection phase, and  
11 measurement invariance during the analyses phase of the data. Below, we expand on the  
12 various issues of concern here.  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22

23 Pooled-survey studies simply assume that samples are invariant, but the issue is not actually  
24 studied. However, recently Rendall *et al.* (2013) have raised concerns about this assumption.  
25 Different ways to deal with a lack of invariance already exist in single country study contexts.  
26 One way to deal with the contextual survey differences includes hierarchical modeling and  
27 parameterization (Gelman *et al.*, 1998; Tighe *et al.*, 2010). This approach is, however, not  
28 appropriate for cross-survey cases with data a small number of countries, as the method  
29 depends on pooling a sufficient number of surveys (e.g. in Gelman *et al.* (1998) the number of  
30 surveys used in the parameterization was 51) to allow for a parameterization of model  
31 parameter variability across the studies (Rendall *et al.*, 2013). This problem has been  
32 addressed by Schenker *et al.* (2010) but as pointed out by Rendall *et al.* (2013) the proposed  
33 method of subdividing surveys' samples into subgroups is also problematic. Rendall *et al.*  
34 (2013) propose a pooled cross-survey MI method, which is preceded by a model-fitting  
35 approach to evaluate the reasonableness of the assumption that surveys are independent  
36 realizations of the same population. However, whether or not the issue of a limited number of  
37 cases at the country level can be addressed, the methods suggested by Rendall *et al.* (2013)  
38 and Schenker *et al.* (2010) can provide only limited help to international marketing  
39 researchers. Thus, international marketing scholars should pay special attention to sampling  
40 invariance, and make sure that if a cross-survey cross-country design is applied, the country  
41 samples should represent the same population.  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 It seems to be a common practice in single group cross-survey research to assume that  
4 survey instruments are invariant. Such an assumption is often considered reasonable in  
5 single-country research settings (although see Rendall *et al.* (2013), for a divergent view).  
6  
7 However, recent literature has faced challenges while trying to account for potential violations  
8 of the assumption of invariance into the multiple imputation analyses (see e.g. Reiter *et al.*,  
9 2006; von Hippel, 2007; Schenker *et al.*, 2010). Even so, whether or not it is considered safe  
10 to ignore the issue of invariance within a single country, it is undeniable that one cannot ignore  
11 it in cross-national research settings.  
12  
13  
14  
15  
16  
17

18  
19 A key issue in this context is measurement invariance, which refers to “whether or not, under  
20 different conditions of observing and studying phenomena, measurement operations yield  
21 measures of the same attribute” (Horn and McArdle, 1992, p. 117). While measurement  
22 invariance is vital to drawing justifiable conclusions from cross-national studies (Steenkamp  
23 and Baumgartner, 1998), it is also a relevant issue for many other types of research.  
24 Specifically, whenever the researcher has reason to suspect that the various groups in a  
25 sample may exhibit variance, they need to consider testing for measurement invariance across  
26 the groups. In various contexts, this could include even single-country studies that incorporate  
27 multiple groups which may exhibit some variance – such as gender, education, hierarchical  
28 position in a firm (e.g. managers versus subordinates) and the like. As such, which the  
29 specifics of our work refer to the cross-national context common to international marketing and  
30 business research, the general issues we raise are relevant to many other research contexts,  
31 and researchers should take care in any possible multi-group situation.  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43

44 Various types of invariance are testable, including configural, metric, scalar, factor variance,  
45 factor covariance, error covariance, and error variance. It is important to note that not all types  
46 of invariance are necessary at all times, depending on the purpose of the researcher  
47 (Steenkamp and Baumgartner, 1998)<sup>3</sup>. However, equally important is that invariance of some  
48 type *is* always necessary for robust cross-national comparisons. In particular, we note that the  
49 combining (i.e. pooling) of multiple single-country data sets into a single (larger) data set  
50  
51  
52  
53  
54

55  
56 <sup>3</sup> It is beyond the scope of this paper to provide detailed discussion of the types of invariance that are necessary for  
57 different research objectives. Researchers are encouraged to explore the canonical work of authors such as Steenkamp and  
58 Baumgartner (1998) for such purposes.  
59  
60

1  
2  
3 requires some of the most stringent invariance testing. This is an important notion for cross-  
4 survey studies, where the imputation is currently often done on a combined data set. However,  
5 as De Jong et al. (2007) note obtaining cross-cultural invariance is often challenging, and  
6 researchers need often accept partial invariances (Steenkamp and Baumgartner, 1998). The  
7 further from 'full' invariance one gets, the more cautious researchers need to be regarding the  
8 strength of the conclusions they can draw across countries, although this is rarely explicitly  
9 addressed in research. Unfortunately, partial invariance is often the best we can achieve. On  
10 the other hand, cross-survey designs can increase the likelihood of full invariance, because  
11 they have the potential to allow greater sample sizes, and greater questionnaire lengths.  
12  
13  
14  
15  
16  
17  
18  
19

20  
21 Given the challenges discussed above, this paper focuses on exploring the performance of  
22 different approaches using multiple imputation procedure for cross-country cross-survey data.  
23 By different approaches it is referred to a usage of different input data in multiple imputations.  
24 In the first case (Model A), the imputation is based on all the data available for the  
25 measurement model of five latent variables. In the second case (Model B), the multiple  
26 imputation is based on a more coherent set of items based on the invariance structure of the  
27 joint items shared across the two groups (i.e. countries here). Whereas the third case (Model  
28 C) is based on even more limited approach, and the items used in multiple imputation are  
29 based both on examination of the invariance structures of the joint items and the performance  
30 of the measurement model in the group where the full data was originally available. Several  
31 different outcomes are then assessed in order to determine which of the approaches would  
32 provide the most reliable imputed data: comparison of imputed means and standard  
33 deviations, proportional differences (the bias) in means and standard deviations, bias in  
34 distributions, standardized loadings in CFA, correlations of latent variables  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

## 50 **An illustrative application of multiple imputation to cross-country research**

### 51 *Overview of the study*

52  
53  
54  
55 To demonstrate our method, we use a two-country data set of constructs relating to  
56 international entrepreneurial-orientation (IEO, e.g. Sundqvist *et al.*, 2012). The choice of IEO  
57  
58  
59  
60



1  
2  
3 as our focal issue is driven simply by data availability, and thus, we do not necessarily aim to  
4 make any contribution to the EO field specifically. As such, significant discussion of conceptual  
5 and operational issues regarding EO is beyond our scope, as is a full review of prior EO work.  
6  
7 Nevertheless, it is useful to know that EO has some limited history of cross-national research.  
8  
9 One of the first tests in this regard was conducted by Knight (1997) between English and  
10 French speaking Canadian managers, using an eight-item scale based on Covin and Slevin  
11 (1989). While invariance was not explored, Knight (1997) did find that the scale performed well  
12 in both English and French with regard to consistency and pattern of factor structure, internal  
13 consistency, convergent and discriminant validity. Similarly, Kreiser et al. (2002) tested a  
14 model of entrepreneurial behavior in six countries, also ignoring invariance issues. More recent  
15 examples where invariance is explored include Hansen et al. (2011), who utilized a seven-  
16 country sample to test the invariance of the Covin and Slevin (1989) EO scale. Unfortunately,  
17 invariance was difficult to reach, and the only pair in which full invariance was achieved was  
18 the US and Greece. Such results suggest strongly that knowledge of cross-country invariance  
19 is very limited in this context.  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31

### 32 *The Measures and data used*

33  
34  
35 We use a multi-country sample of real data as an example of cross-sample imputation,  
36 consisting of 292 exporting firms from New Zealand and 302 from Finland. It is important to  
37 note that our use of a single empirical example is intended as an illustration of the issues  
38 involved in cross-national / cross-survey imputation, rather than as an authoritative  
39 examination of the extent to which variance induces bias, and the different methods of dealing  
40 with it (which would require a simulation study). We used a set of constructs designed to  
41 capture concepts related to IEO. Autonomy (AUTO) and competitive aggressiveness (COMP)  
42 were measured with 7-point Likert scales, varying from complete disagreement to complete  
43 agreement. Autonomy consisted of four items based on Jambulingam et al.'s (2005) autonomy  
44 scale items, adapted to the international context. Competitive aggressiveness was also  
45 measured with four items based on items from Narver and Slater's (1990) competitor-  
46 orientation scale, and Jaworski and Kohli's (1993) market responsiveness scale.  
47  
48 Innovativeness (INNO), proactiveness (PRO) and risk taking (RISK) were measured with 9-  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 point Likert scales, varying from complete disagreement to complete agreement. Our  
4 innovativeness measure was based on the scale of Jambulingam et al. (2005). We used  
5 Jambulingam et al.'s (2005) proactiveness scale extended to identify which managers seized  
6 the opportunities in the anticipation of future market conditions. Proactiveness was  
7 represented with three items. Risk taking was measured with three items drawn from  
8 Jambulingam et al.'s (2005) risk-taking scale. All scale items were modified to capture the  
9 international aspects of the concept of interest.  
10  
11  
12  
13  
14  
15  
16  
17

18 Originally the datasets included complete information on all variables under study, but were  
19 subsequently modified in order to demonstrate our imputation procedure. For creating the  
20 hypothetical situation of completely missing variables in one country, the items representing  
21 competitive aggressiveness and autonomy were deleted from the New Zealand data. The  
22 Finnish data with 302 responses included full data for all the variables.  
23  
24  
25  
26  
27  
28  
29  
30

### 31 *Imputation procedure and data preparation*

32  
33 The multiple imputation procedure was conducted using SAS 9.3 software. The imputation  
34 procedure followed the Markov Chain Monte Carlo (MCMC) method (see Schafer, 1997).  
35 MCMC has been found to provide valid imputed values for cross-sectional missing data  
36 (Schunk, 2006; Shrive et al., 2006; Ziegelmeyer, 2013). Following Honaker and King (2010)  
37 the first step is to bootstrap the concatenated data set to create  $m$  versions of the incomplete  
38 data, where  $m$  ranges typically from 3 to 5 as in other multiple imputation approaches.  
39 Bootstrap resampling involves taking a sample of size  $n$  with replacement from the original  
40 dataset. Here, the  $m$  bootstrap samples of size  $n$  are obtained from the concatenated file,  
41 where  $n$  is the total sample size of the file. Second, for each bootstrapped data set, the EM  
42 (Expectation Maximization) algorithm is run. The further analyses also apply LISREL software  
43 for invariance testing and confirmatory factor analyses.  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54

55 Before imputing the missing values, careful look was taken at the descriptive information of the  
56 variables and their histograms. The two items related to autonomy were heavily skewed on the  
57  
58  
59  
60



1  
2  
3 right and therefore the distributions were corrected by taking a logarithm of the items, because  
4 the chosen imputation method tends to correct the distribution towards normality (see Schafer,  
5 1997; Schafer and Olsen, 1998). Additional specifications were made beforehand to the  
6 imputation procedure. As competitive aggressiveness was measured with a seven-point Likert  
7 scale, the requested specifications included a minimum of one and maximum of seven. In  
8 addition, a request was made for zero rounding. Both these specifications ensure the imputed  
9 data conforms to the existing data structure. Similarly, the autonomy items that were log  
10 transformed, had their minimum set at zero ( $\ln 1$ ) and maximum fixed to 1.95 ( $\ln 7$ ). For the  
11 imputation process only those variables were used that were included in the measurement  
12 model. Running the MCMC method for multiple imputation produced five imputed datasets.  
13 Because the logarithm was used for the autonomy variables, transformation to original scale  
14 was conducted before further analysis of the results. This was done with exponent function  
15 and rounding the values to the whole number.  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28

29 *Model A: Imputing all missing variables for New Zealand with all measurement items*

30  
31 A hypothetical situation is created in a following manner. A researcher has collected good data  
32 from Finland, including four items to measure competitive aggressiveness and four items  
33 reflecting autonomy. Additionally, concepts of innovativeness, proactiveness and risk taking  
34 each have three measurement items. Another researcher has collected data from New  
35 Zealand, but only measuring innovativeness, proactiveness, and risk taking (using the same  
36 items as in Finland), and missing out competitive aggressiveness and autonomy by design  
37 (perhaps to shorten the questionnaire). For joining the datasets (for instance for comparative  
38 purposes), imputation is used as tool for estimating values for the missing items in New  
39 Zealand data. The starting point for this type of case is illustrated in Figure 1.  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20

	COMP (4 var.)				AUT (4 var.)				INNO (3 var.)			PROACT (3 var.)			RISK (3 var.)					
F I																				
N Z																				

21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56

Figure 1. Missingness for New Zealand data in Model A

25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56

The MCMC imputation procedure produced, as requested, five imputed datasets. The correction related to the skewness of items reflecting autonomy was taken into account beforehand. For the Finnish data, a logarithmic transform was computed and these variables were used in the imputation. Afterwards, the original scale was retrieved using the exponent function and rounding the values to match the original scale. The imputed data was evaluated by comparing the imputed means and standard deviations obtained with imputation to the indicators computed from the original New Zealand data. Here, it must be remembered that we removed the New Zealand data for the purposes of this imputation, and are in fact in possession therefore of the 'real' data on competitive aggressiveness, and autonomy from New Zealand. The overall imputed mean and standard deviation were computed from all the imputed data. A proportional difference is used for illustrating the "success" of imputation. As can be seen, the differences in the means of items measuring competitive aggressiveness were mainly less than five percent. However, the difference in the standard deviations was not very consistent. This is also the case for autonomy. The difference in standard deviation varies, and also there is variation in the mean differences (see Table 1).

57  
58  
59  
60

Table 1. Imputed means and standard deviations for Model A

<i>Item</i>		Original FIN	Original NZ	Imputed data sets for NZ					Overall imputed mean / std. dev.	Proportional difference
				1	2	3	4	5		
COMP1	Mean	4.934	4.856	4.901	4.678	4.682	4.901	4.582	4.749	.022
	Std. dev.	1.534	1.569	1.303	1.431	1.354	1.405	1.473	1.398	.109
COMP2	Mean	4.633	4.592	4.469	4.500	4.644	4.565	4.377	4.511	.018
	Std. dev.	1.472	1.425	1.439	1.288	1.343	1.483	1.408	1.395	.021
COMP3	Mean	5.599	5.497	5.274	5.271	5.438	5.404	5.267	5.331	.030
	Std. dev.	1.339	1.386	1.167	1.148	1.178	1.134	1.220	1.171	.155
COMP4	Mean	4.870	4.952	4.688	4.565	4.719	4.658	4.589	4.644	.062
	Std. dev.	1.312	1.333	1.182	1.184	1.234	1.273	1.383	1.253	.060
AUTO1	Mean	2.114	2.825	2.428	2.548	2.661	2.404	2.353	2.479	.123
	Std. dev.	1.214	1.334	1.136	1.244	1.243	1.155	1.119	1.184	.112
AUTO2	Mean	1.937	2.339	2.257	2.253	2.421	2.288	2.229	2.290	.021
	Std. dev.	1.381	1.356	.973	1.165	1.203	1.190	1.068	1.124	.172
AUTO3	Mean	1.994	2.212	2.212	2.336	2.339	2.236	2.257	2.276	-.029
	Std. dev.	1.249	1.348	1.034	1.102	1.077	.993	1.051	1.052	.220
AUTO4	Mean	2.599	2.736	2.812	2.925	2.815	2.952	2.853	2.871	-.049
	Std. dev.	1.495	1.460	1.314	1.365	1.290	1.369	1.325	1.332	.088
<b>Overall mean for proportional difference</b>									Mean	.044
									Std. dev.	.117

Table 2. Testing the equality of original and imputed distributions in Model A

<i>Item</i>	<i>KSa</i>	<i>p</i>
COMP1	1.325	<.100
COMP2	0.652	>.100
COMP3	2.618	<.001
COMP4	2.073	<.001
AUTO1	1.859	<.010
AUTO2	1.004	>.100
AUTO3	1.848	<.010
AUTO4	1.539	<.050

KSa = asymptotic Kolmogorov-Smirnov statistic

For additional precision, the original and imputed distributions of the New Zealand data were compared with a Kolmogorov-Smirnov test for equal distributions (Gibbons and Chakraborti, 1992). Table 2 presents the results of analysis. As can be seen, only two items (comp2 and auto3) provided similar distribution between imputed and original data.

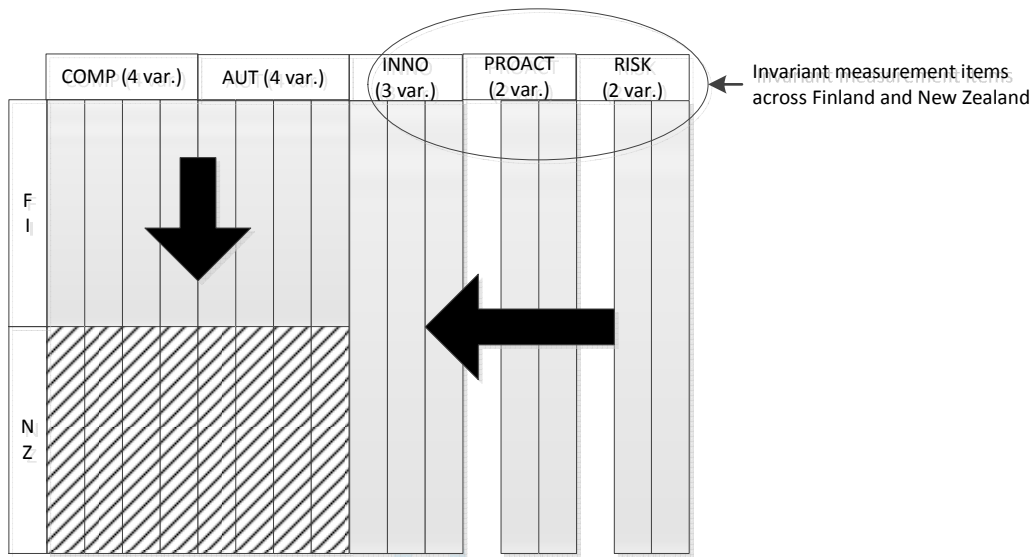
*Model B: Imputation for New Zealand based on invariant joint measurement items*

As could be seen from the comparison results of Model A, the imputation performed purely based on generally accepted measures was not very accurate. In order to get more precise estimates for imputed values, the three concepts common to the data sets (innovativeness, proactiveness, and risk taking) were first analyzed for measurement invariance. Each of the three concepts was assessed for configural invariance, metric invariance and factor variance invariance, with the results presented in Table 3. The steps of invariance were analyzed using the deterioration of the model in terms of changes in chi-square (e.g. Jöreskog, 1971). The results indicate that some deterioration occurred during the invariance analysis, but the changes in chi-square were insignificant.

Table 3. Invariance analysis across countries with three concepts

<i>Level of invariance</i>	<i><math>\chi^2</math> (df)</i>	<i><math>\Delta\chi^2</math> (<math>\Delta</math>df)</i>	<i>RMSEA</i>	<i>CFI</i>
Configural invariance	55.15 (22)	-	.070	.991
Metric invariance	62.89 (26)	7.74 (4)	.068	.990
Factor variance invariance	66.45 (29)	3.56 (3)	.064	.989

The invariance analysis of innovativeness, proactiveness and risk taking revealed that all the items measuring innovativeness were invariant across countries while proactiveness as well as risk taking were invariant with two measurement items for each concept. Therefore, the next demonstration of imputation is based only on the items that reached measurement invariance (Figure 2). In other words, we replicate the situation where the researcher tests for measurement invariance amongst the common construct measures *before* imputing data across countries for the other measures of latent constructs.



22 Figure 2. Missingness for New Zealand data in Model B

23  
24  
25  
26  
27 Using three items of innovativeness, two items of proactiveness and two of risk taking, the  
28 multiple imputation was conducted in similar manner as presented above. Five data sets were  
29 created and the results were analyzed in terms of means, standard deviations and equality of  
30 distributions. The means of the imputed items were on the same level as in Model A (Table 4).  
31 The standard deviations however were more divergent from original data than in the previous  
32 model. Although the distribution of items covering autonomy were transformed with logarithm  
33 before imputation, the imputation failed for the first autonomy item and the estimated mean  
34 had a high proportional difference compared to the original mean. That said, the distributions  
35 resulting after imputation fit better to the original data (Table 5). One item from competitive  
36 aggressiveness and two items from autonomy have similar distributions after imputation  
37 compared to the original New Zealand data.  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Table 4. Imputed means and standard deviations for Model B

<i>Item</i>		Original FIN	Original NZ	Imputed data sets for NZ					Overall imputed mean /std. dev.	Proportional difference
				1	2	3	4	5		
COMP1	Mean	4.934	4.856	4.795	4.688	4.668	4.678	4.582	4.682	.036
	Std. dev.	1.534	1.569	1.372	1.432	1.458	1.347	1.384	1.399	.108
COMP2	Mean	4.633	4.592	4.435	4.250	4.425	4.466	4.507	4.416	.038
	Std. dev.	1.472	1.425	1.411	1.343	1.346	1.214	1.264	1.319	.074
COMP3	Mean	5.599	5.497	5.346	5.212	5.284	5.233	5.534	5.322	.032
	Std. dev.	1.339	1.386	1.149	1.256	1.174	1.235	1.173	1.202	.133
COMP4	Mean	4.870	4.952	4.640	4.658	4.743	4.521	4.682	4.649	.061
	Std. dev.	1.312	1.333	1.197	1.281	1.127	1.188	1.313	1.224	.082
AUTO1	Mean	2.114	2.825	2.408	2.315	2.438	2.291	2.373	2.365	.163
	Std. dev.	1.214	1.334	1.066	1.034	1.152	.995	1.100	1.071	.198
AUTO2	Mean	1.937	2.339	2.435	2.205	2.384	2.120	2.192	2.267	.031
	Std. dev.	1.381	1.356	1.227	1.106	1.156	.975	.951	1.093	.194
AUTO3	Mean	1.994	2.212	2.291	2.233	2.318	2.045	2.271	2.232	-.009
	Std. dev.	1.249	1.348	1.091	1.026	1.089	.890	1.048	1.035	.233
AUTO4	Mean	2.599	2.736	2.771	2.777	2.832	2.682	2.777	2.768	-.012
	Std. dev.	1.495	1.460	1.379	1.230	1.293	1.248	1.285	1.287	.118
<b>Overall mean for proportional difference</b>								Mean	.048	
								Std. dev.	.143	

Table 5. Testing the equality of original and imputed distributions in Model B

<i>Item</i>	<i>KSa</i>	<i>p</i>
COMP1	1.624	<.050
COMP2	1.165	>.100
COMP3	2.607	<.001
COMP4	2.319	<.001
AUTO1	2.489	<.001
AUTO2	1.047	>.100
AUTO3	1.784	<.001
AUTO4	1.079	>.100

KSa = asymptotic Kolmogorov-Smirnov statistic

*Model C: Imputation for New Zealand after completing the measurement model with Finnish data*

Based on the two trials presented above, there still remains a need to improve the estimates provided with imputation. As such, for a third trial, we developed the full measurement model for five concepts using the Finnish data, with confirmatory factor analysis, as well as only using the invariant items for the common latent construct measures. As a result, a number of items were dropped from the measurement model. Two items remained in the measure of competitive aggressiveness and two in the measure of autonomy. Changes were also made for proactiveness and risk taking; for both, one item was dropped from the measurement model. Considering the goodness of fit statistics for the model ( $\chi^2=57.38$  ( $p=.010$ ),  $df=34$ ,  $RMSEA=.046$ ,  $GFI=.995$ ,  $AGFI=.990$ ,  $NFI=.979$ ,  $NNFI=.986$ ), it can be concluded that the measurement model has a rather good fit to the data (see e.g. Hair *et al.*, 1998; Hayduk, 1989; Kelloway, 1998). The results of the model also suggested a good level of reliability, as the composite reliability coefficient exceeded the critical value of .70 and coefficient describing average variance extracted exceeds the value of .50 for all the concepts (e.g. Fornell and Larcker, 1981). The last imputation model is illustrated in Figure 3.

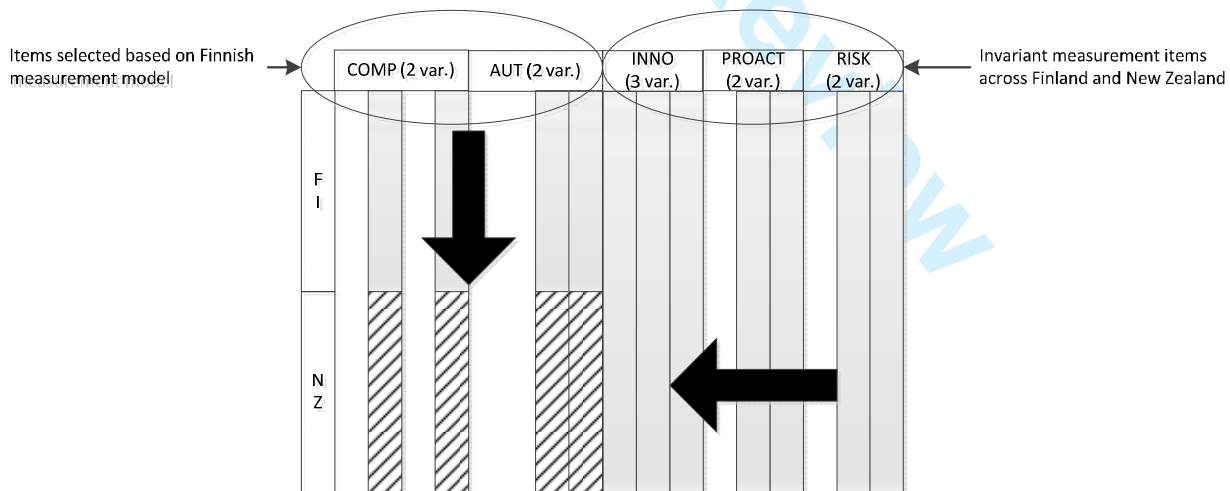


Figure 3. Missingness for New Zealand data in Model C

Table 6 includes the comparison of imputed means and standard deviations. This model succeeded rather well concerning the item means. The proportional difference for all items was less than five percent. Additionally, the difference related standard deviations is also much smaller than in the previous two models. However, there is still a lot of error in the standard deviations related to items measuring autonomy.

Table 6. Imputed means and standard deviations for Model C

Item		Original FIN	Original NZ	Imputed datasets for NZ					Overall imputed mean	Proportional mean difference
				1	2	3	4	5		
COMP2	Mean	4.633	4.592	4.377	4.596	4.747	4.449	4.699	4.573	.004
	Std. Dev.	1.472	1.425	1.41	1.353	1.389	1.448	1.349	1.390	.025
COMP4	Mean	4.87	4.952	4.565	4.736	4.771	4.764	4.757	4.718	.047
	Std. Dev.	1.312	1.333	1.243	1.202	1.226	1.312	1.232	1.243	.068
AUTO4	Mean	1.994	2.212	2.147	1.993	2.127	2.144	2.298	2.142	.032
	Std. Dev.	1.249	1.348	.999	.949	.985	.981	1.057	.994	.262
AUTO5	Mean	2.599	2.736	2.726	2.788	2.664	2.805	2.764	2.749	-.005
	Std. Dev.	1.495	1.46	1.342	1.366	1.197	1.337	1.275	1.303	.107
<b>Overall mean for proportional difference</b>									Mean	.022
									Std. dev.	.115

Finally, the evaluation related to the equality of distributions is promising. This model retrieved distribution equality for three out of four items. Only the second item of competitive aggressiveness had a distribution that differed from original.

Table 7. Testing the equality of original and imputed distributions in Model C

Item	KSa	p
COMP2	0.363	>.100
COMP4	1.955	<.050
AUTO3	1.100	>.100
AUTO4	1.026	>.100

KSa = asymptotic Kolmogorov-Smirnov statistic



Figure 4 presents the distributions after imputation and the original distributions. The problem related to variable comp4, is clearly attached to the original distribution. The histogram illustrates the tendency that the imputation method has related to the assumption of normality. Therefore the imputed values are “corrected” toward normality as a consequence of imputation.

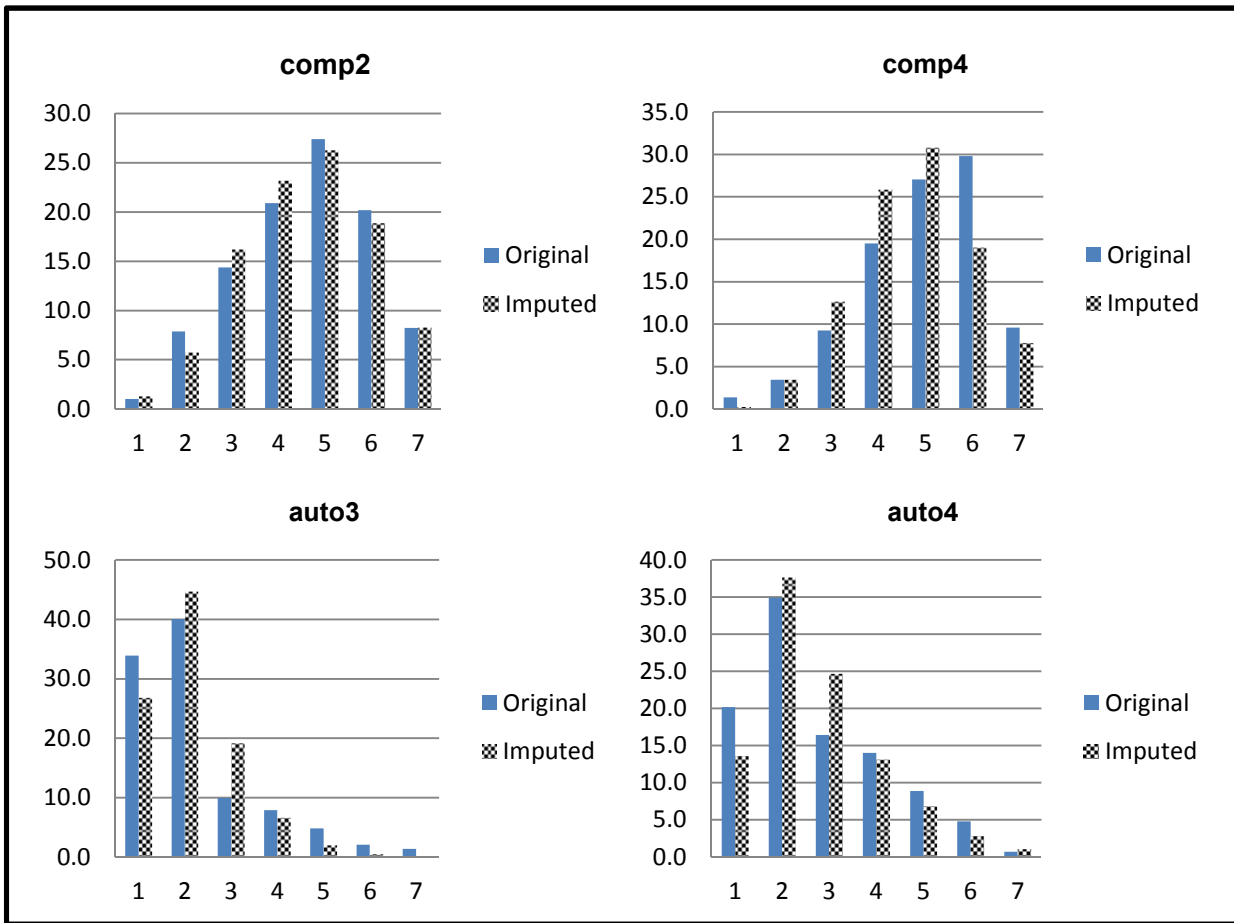


Figure 4. Distribution of imputed variables vs original variables

However, all in all Model C seems to provide most accurate imputation, based on distribution comparisons.

### **Comparison of measurement models and inferences about latent constructs**

Next, we compared the three imputation models (Models A, B, and C) in terms of how the measurement model of latent constructs corresponded with the original full data results in New Zealand. For these comparisons we used a confirmatory factor analysis model with 'best' measurement invariance that we could reach with original full data sets.

This same invariant CFA model was estimated for Models A, B and C. The procedure for these analyses was the following:

1. Calculation of measurement item means, variances and covariances for all five imputed datasets for Models A, B, and C.
2. Combination of means, variances and covariances of the five imputed datasets for Models A, B, and C. For this combination, we averaged the five estimates into single estimate with arithmetic mean.
3. Conducting multi-group CFA for Finnish and NZ imputed (and combined) data with the same model specifications than with original full data.

Table 8. Standardized loadings and correlations

Loadings	Model A		Model B		Model C		Original	
	Comp	Auto	Comp	Auto	Comp	Auto	Comp	Auto
COMP2	0.635		0.619		0.646		0.618	
COMP4	0.874		0.887		0.855		0.911	
AUTO3		0.820		0.684		0.677		0.967
AUTO4		0.681		0.814		0.838		0.698
Correlations								
	Comp	Auto	Comp	Auto	Comp	Auto	Comp	Auto
Comp	1.000		1.000		1.000		1.000	
Auto	-0.050	1.000	0.006	1.000	-0.114	1.000	-0.041	1.000
Pro	0.471	-0.196	0.508	-0.034	0.544	-0.186	0.542	-0.233
Inno	0.426	-0.105	0.479	0.002	0.469	-0.135	0.538	-0.238
Risk	0.233	-0.026	0.271	0.015	0.230	-0.068	0.333	-0.128

The standardized loadings for measurement items and correlations between latent constructs are presented in Table 8. We can see some differences in the standardized estimates between different imputation models. The measure loadings for Comp –items are relatively similar between imputation models and Original full data, but for Auto the Model A seems to outperform Models B and C. We can observe differences in the correlations as well.

In order to explore which of the imputation models' standardized loadings and correlations are closest to original full data estimates, we calculated deviances (see Table 9). These deviances were calculated by subtracting the estimate of full original data from estimate of imputed data (e.g. deviance for Model A correlation between Auto and Comp is  $-0.050 - (-0.041) = -0.009$ ). In Table 9, the smallest deviances are bolded. It seems that in terms of measure loadings, the Model A has the closest estimates for Auto and Model B for Comp. In terms of correlations the results are more mixed, although Model C has highest number of closest estimates.

Table 9. Deviations of loadings and correlations between original full data and imputed data.

Loadings	Model A		Model B		Model C	
	Comp	Auto	Comp	Auto	Comp	Auto
COMP2	0.017		<b>0.001</b>		0.028	
COMP4	-0.037		<b>-0.024</b>		-0.056	
AUTO3		<b>-0.147</b>		-0.283		-0.290
AUTO4		<b>-0.017</b>		0.116		0.140
Correlations	Comp	Auto	Comp	Auto	Comp	Auto
Comp	0		0		0	
Auto	<b>-0.009</b>	0	0.047	0	-0.073	0
Pro	-0.071	<b>0.037</b>	-0.034	0.199	<b>0.002</b>	0.047
Inno	-0.112	0.133	<b>-0.059</b>	0.240	-0.069	<b>0.103</b>
Risk	-0.100	0.102	<b>-0.062</b>	0.143	-0.103	<b>0.060</b>

Smallest deviations are bolded

In order to explore how the choices of imputation method influences the statistical inferences about measurement of latent constructs and their covariances, we also made similar comparisons for unstandardized estimates and their standard errors and corresponding t-

values (see Table 10). The estimated loadings of all Models A, B, and C were significant at  $p = 0.05$  (cut-off limit 1.96), which corresponds with the results of original full New Zealand data.

Table 10. Unstandardized loading estimates.

Items	Model A		Model B		Model C		Original data	
	Comp	Auto	Comp	Auto	Comp	Auto	Comp	Auto
COMP2	1.000		1.000		1.000		1.000	
COMP4	1.236 (0.133) <i>9.325</i>		1.302 (0.138) <i>9.424</i>		1.184 (0.118) <i>10.033</i>		1.292 (0.184) <i>7.026</i>	
AUTO3		1.000		1.000		1.000		1.000
AUTO4		1.022 (0.247) <i>4.131</i>		1.449 (0.439) <i>3.302</i>		1.547 (0.393) <i>3.932</i>		0.847 (0.204) <i>4.161</i>

Standard errors in parentheses; t-values in *italics*

Again, we also calculated the deviance between imputed data models and original full New Zealand data model (see Table 11). As was shown with standardized loadings before, Model A provided closest estimates for Auto and Model B for Comp, whereas the deviations were largest in Model C.

Table 11. Deviations from Original full data loadings

Items	Model A		Model B		Model C	
	Comp	Auto	Comp	Auto	Comp	Auto
COMP4	-0.056 (0.051) <b>2.299</b>		<b>0.01</b> (0.046) <i>2.398</i>		-0.108 (0.066) <i>3.007</i>	
AUTO4		<b>0.175</b> (-0.043) <b>-0.030</b>		0.602 (-0.235) <b>-0.859</b>		0.700 (-0.189) <b>-0.229</b>

Standard errors in parentheses; t-values in *italics*; Smallest deviations bolded

The same comparison was done for covariances between latent constructs and corresponding standard errors and t-values (Table 12). In these comparisons differences in the statistical inferences regarding covariances can be observed. The covariance between Auto and Pro is negative and significant ( $p = 0.05$ , cut-off limit 1.96) in the original New Zealand data, but Model B fails to reject the null hypothesis. Also the covariance between Auto and Inno is negative and significant at  $p = 0.05$  in the original data, but all of the Models fail to reject null hypothesis at  $p = 0.05$ . However, Model C is the closest since the negative covariance is significant at  $p = 0.10$  (cut-off limit 1.65). The covariance between Auto and Risk is negative and significant ( $p = 0.05$ ) in the original full data, but all Models fail to reject null hypothesis.

Table 12. Estimated covariances between latent variables

PHI	Model A		Model B		Model C		Original data	
	Comp	Auto	Comp	Auto	Comp	Auto	Comp	Auto
Comp	0.834 (0.124) <i>6.729</i>		0.753 (0.113) <i>6.666</i>		0.856 (0.122) <i>7.047</i>		1.573 (0.331) <i>4.754</i>	
Auto	-0.044 (0.067) <i>-0.657</i>	0.917 (0.231) <i>3.975</i>	0.004 (0.053) <i>0.073</i>	0.628 (0.199) <i>3.150</i>	-0.081 (0.057) <i>-1.414</i>	0.590 (0.160) <i>3.681</i>	-0.089 (0.145) <i>-0.614</i>	3.086 (0.768) <i>4.016</i>
Pro	0.608 (0.104) <i>5.858</i>	-0.266 (0.099) <i>-2.675</i>	0.624 (0.100) <i>6.220</i>	-0.038 (0.083) <i>-0.462</i>	0.709 (0.106) <i>6.719</i>	-0.202 (0.085) <i>-2.367</i>	1.805 (0.345) <i>5.226</i>	-1.087 (0.308) <i>-3.525</i>
Inno	0.600 (0.108) <i>5.566</i>	-0.155 (0.104) <i>-1.492</i>	0.641 (0.105) <i>6.097</i>	0.003 (0.087) <i>0.032</i>	0.669 (0.109) <i>6.136</i>	-0.160 (0.087) <i>-1.828</i>	2.480 (0.470) <i>5.279</i>	-1.535 (0.418) <i>-3.670</i>
Risk	0.367 (0.113) <i>3.256</i>	-0.043 (0.120) <i>-0.363</i>	0.408 (0.107) <i>3.793</i>	0.020 (0.101) <i>0.199</i>	0.370 (0.115) <i>3.226</i>	-0.090 (0.097) <i>-0.932</i>	1.440 (0.358) <i>4.019</i>	-0.778 (0.392) <i>-1.985</i>

Standard errors in parentheses; t-values in *italics*

Again, we calculated deviances from original full New Zealand data for results of different models (see Table 13). These deviances were calculated by subtracting the estimate of full original data from estimate of imputed data (e.g. deviance for Model A covariance between Auto and Comp is  $-0.044 - (-0.089) = 0.045$ ). In terms of covariance estimates, the Model C has 6 closest estimates (Model A has 2 closest estimates and Model B has 1 closest

estimate). The closest standard errors are more evenly distributed between Model A (5 closest estimates) and Model C (4 closest estimates), whereas Model B has only one closest estimate, which is actually of same magnitude with Model C. Model A has 5 closest estimates of t-values, whereas Models B and C have both two closest t-value estimates.

Table 13. Deviations from original full data covariances

PHI	Model A		Model B		Model C	
	Comp	Auto	Comp	Auto	Comp	Auto
Comp	-0.739 <b>(0.207)</b> <i>1.975</i>		-0.820 <i>(0.218)</i> <b>1.912</b>		<b>-0.717</b> <i>(0.209)</i> 2.293	
Auto	0.045 <b>(0.078)</b> <i>-0.043</i>	<b>-2.169</b> <b>(0.537)</b> <i>-0.041</i>	0.093 <i>(0.092)</i> 0.687	-2.458 <i>(0.569)</i> <i>-0.866</i>	<b>0.008</b> <i>(0.088)</i> <i>-0.800</i>	-2.496 <i>(0.608)</i> <i>-0.335</i>
Pro	-1.197 <i>(0.241)</i> <b>0.632</b>	<b>0.821</b> <b>(0.209)</b> <b>0.850</b>	-1.181 <i>(0.245)</i> <i>0.994</i>	1.049 <i>(0.225)</i> <i>3.063</i>	<b>-1.096</b> <b>(0.239)</b> <i>1.493</i>	0.885 <i>(0.223)</i> <i>1.158</i>
Inno	-1.880 <i>(0.362)</i> <b>0.287</b>	1.380 <i>(0.314)</i> 2.178	-1.839 <i>(0.365)</i> <i>0.818</i>	1.538 <b>(0.331)</b> <i>3.702</i>	<b>-1.811</b> <b>(0.361)</b> <i>0.857</i>	<b>1.375</b> <b>(0.331)</b> <b>1.842</b>
Risk	-1.073 <i>(0.245)</i> <i>-0.763</i>	0.735 <b>(0.272)</b> 1.622	<b>-1.032</b> <i>(0.251)</i> <b>-0.226</b>	0.798 <i>(0.291)</i> 2.184	-1.070 <b>(0.243)</b> <i>-0.793</i>	<b>0.688</b> <i>(0.295)</i> <b>1.053</b>

Standard errors in parentheses; t-values in *italics*; Smallest deviations are bolded

Overall, it seems that the results obtained by using Model C are closest to the results with original full New Zealand data. The accuracy of the measurement loading estimates was not the best with this imputation method, but the estimated covariances between latent constructs were the most accurate with Model C, which lead also to most accurate inferences about statistical significance of these covariances. However, although Model C provides the most promising results, none of the imputation models leads to the same statistical inferences about covariances between latent constructs as the original full data.

## Discussion

The benefits of cross-survey MI are potentially very large (Rendall *et al.*, 2013). Split questionnaire designs and multiple imputation techniques provide enormous possibilities for researchers, and support development of cross-country research networks. Currently it is obvious that survey based research is conducted globally in different research fields, even by researchers with similar interests (e.g. strategy, marketing, and information systems researchers often examine similar issues in slightly differing ways). Conducting cross-country comparisons by merging the collected data, and complementing the missing elements with cross-survey MI theoretically makes it possible to maximize the output and effort placed on data collection. In some countries where extensive mail surveys tend not gain a good response rate it could be possible, for example, to collect full questionnaires via interviews from a smaller pool of respondents and then use split questionnaires and MI to gain a larger  $N$ . However, based on our study (albeit a single exploratory example), there still exist several serious issues that need to be examined before such an approach can be recommended. While it is impossible to draw authoritative conclusions without the use of a simulation study, the findings of our research still provide important first steps towards a greater understanding of the impact of cross-group measurement variance in situations where imputation is needed. Such issues are highly relevant not only to international marketing, but also to many other social science contexts. Indeed, we agree that accounting for differences in survey sampling and measurement characteristics across the multiple surveys is a considerable barrier to the successful implementation of cross-survey MI in the social sciences (Rendall *et al.*, 2013).

In this paper we explored the performance of MI in the type of cross-country cross-survey setting often relevant for international marketing research. As a starting point we had complete data on 5 constructs measured with 17 measurement items (Comp with 4 items, Aut with 4 items, Inno with 3 items, Pro with 3 items, Risk with 3 items) from one country (Finland), and we had missing data on 2 of these constructs (Comp and Aut), i.e. 8 missing measurement items from another country (New Zealand). This missingness was created by removing data from the New Zealand data set, to simulate a typical 'missing by design' situation, and allow us

1  
2  
3 to compare the various imputation methods with the real data. We explored how the missing  
4 data could be imputed using data from both countries with three different imputation methods.  
5  
6

7  
8 The first imputation method (Model A) used all existing data from both countries to impute the  
9 missing 8 measurement items. The ratios of imputed data to used data is 8/9 within New  
10 Zealand (8 imputed NZ variables, 9 measured NZ variables used for imputation) and 8/17  
11 between New Zealand and Finland (8 imputed NZ variables, 17 measured FIN variables used  
12 for imputation). In this first method, there was no information on the cross-country quality of the  
13 measurement items. The second method (Model B) was based on assessing measure  
14 invariance before imputation. We used only the invariant measurement items of Inno, Pro and  
15 Risk for imputing all of the missing 8 variables. While this method uses between country  
16 knowledge on measurement quality, it simultaneously increases the ratios of imputed data to  
17 used data to 8/7 within New Zealand and 8/15 between Finland and New Zealand.  
18  
19

20  
21 The third method (Model C) was based on reducing the number of imputed variables using  
22 within country information from Finland. Confirmatory factor analysis with Finnish data  
23 revealed that 2 measures qualify for measuring Comp and 2 measurement items for  
24 measuring Aut. Based on this analysis, we decided to impute only these measurement items  
25 into New Zealand data. This analysis enhanced the ratios of imputed data to used data into 4/7  
26 within New Zealand and 4/11 between New Zealand and Finland.  
27  
28

29  
30 The comparison of results revealed that the third method (Model C) provided the most  
31 accurate estimates and inferences about the constructs. Our interpretation of the findings is  
32 that using knowledge on between country measurement qualities may improve the imputation  
33 results. However, this benefit comes with a downside, since it simultaneously reduces the  
34 amount of data used for imputation. The second imputation method (Model B), performed quite  
35 poorly, probably because of the large ratios of imputed data to used data.  
36  
37

38  
39 Even without the use of a simulation study, our results show that cross-country, cross-survey  
40 data has unique challenges, and that there is a likelihood of ending up with erroneous data,  
41 unreliable measures and incorrect conclusions about the relationships between the focal  
42 concepts of interest if these issues are not taken into consideration. The three methods of  
43 dealing with missingness that we explored all resulted in different outcomes, and while each  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 method had certain strengths, none of them could replicate the original data structure  
4 particularly well. This is a highly striking result, which indicates the need for further studies on  
5 the topic.  
6  
7

8  
9  
10 In addition to just comparing the distributions of the imputed data, we also compared the  
11 inferences on covariances between constructs. We examined 7 covariances between latent  
12 constructs, and the baseline results (i.e. those using the original full New Zealand data) were:  
13

- 14 - one non-significant (Auto and Comp),
- 15 - 3 significant positive covariances (Comp with Pro, Inno, and Risk),
- 16 - 3 significant negative (Auto with Pro, Inno, and Risk) covariances.

17  
18  
19  
20 All three imputation methods (Models A, B, and C) led to the same inferences about non-  
21 significant covariance between Auto and Comp, and the 3 significant positive covariances of  
22 Comp with other constructs. However, the significant negative covariances of Auto with other  
23 constructs we rather poorly reproduced with the imputed datasets. The covariance between  
24 Auto and Pro, was found significant with Models A and C, while Model B failed to reject the null  
25 hypothesis. The covariance between Auto and Inno was found clearly non-significant in  
26 Models A and B. Model C provided the most accurate results, but in order to reject the null  
27 hypothesis we would have to accept higher probability of Type I Error ( $p = 0.10$ ). Finally, the  
28 covariance between Auto and Risk was non-significant in all of the imputed models.  
29  
30

31  
32  
33 This finding is notable and highlights that the accuracy of distribution comparisons between  
34 imputed data and full data may not necessarily offer sufficient evidence to make conclusions  
35 about the appropriateness of the imputation methods. Model C provided the most accurate  
36 results both in the comparison of measurement item distributions and in the comparisons of  
37 construct covariances. However, it is interesting to see that the distribution comparisons (for  
38 Model C) showed that distributions of all measurement items for Auto (i.e. AUTO3 and  
39 AUTO4) fitted with original distributions, whereas the distribution of measurement item for  
40 Comp (i.e. COMP4) did not fit with original distribution. Yet, the problems with the inferences  
41 about the latent constructs occurred with Auto construct and not with Comp.  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55

56 *Conclusions, recommendations, and future research*  
57  
58  
59  
60

1  
2  
3 The present study appears to be this first to investigate the issue of measurement variance  
4 when conducting multiple imputation in a cross-group/cross-survey framework. As such, our  
5 study's conclusions should necessarily be taken as exploratory. However, even so, they  
6 highlight several concerns and issues that should be taken into account when planning cross-  
7 country cross-surveys (or more general split-questionnaire or sub-sampling designs), and will  
8 hopefully spur greater attention to be given to this issue in future research within international  
9 marketing and business, and social science more generally. Even if there are several  
10 advantages available for well-implemented cross-country cross-survey (CCCS) designs such  
11 as shorter questionnaires and improved response rates, the concerns pointed out in this paper  
12 lead us to question the appropriateness of the CCCS approach in general, due to the need to  
13 impute across the samples. In particular, the latter is problematic since, in cross-country  
14 studies, the samples may not be justifiably assumed to be drawn from the same population,  
15 and there might also be contextual differences (variances) across the samples, creating biases  
16 in imputation results. Further, if the CCCS approach is not appropriate for cross-country  
17 studies, how should missing data patterns be designed for cross-country studies to take  
18 advantage of the potential of the method in expanding research potentials and statistical power  
19 (c.f. Littvay (2009) for single country solution). While we look at this issue in the specific  
20 context of international marketing, it is highly relevant to all social science situations where  
21 cross-group measurement invariance may be in question.  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36

37  
38 Current practice appears to be to use the imputed data for analysis through a pooling method  
39 proposed by Rubin (1987). In general, if for instance a regression model is created with  
40 imputed data, all the variables in the model are used as a basis for imputation. However, our  
41 results here may question that approach, and for cross-country cases the applicability of such  
42 a method is still uncertain.<sup>4</sup> Further discussion related to measurement reliability after  
43 imputation can be found from van Buuren (2010), who also conducted an imputation  
44 simulation using only items that belonged to the measurement model. In other words, right  
45 now, it is far from certain that researchers have solid advice on exactly which items to use  
46 when imputing in cross-national contexts. It is therefore vital that systematic simulation  
47  
48  
49  
50  
51  
52  
53

---

54  
55 <sup>4</sup> Indeed, while we do not report the results here due to length restrictions, we conducted an additional experiment, using  
56 many other additional variables (e.g. firm descriptives, performance variables) to perform the imputation. Results obtained  
57 from imputation revealed that the imputed means did not correspond well with the original means (results available from  
58 corresponding author on request).  
59  
60

1  
2  
3 experimental research is conducted to provide more generalizable evidence of the best  
4 approach. Such research is the only way of deriving conclusive evidence of the influence of  
5 measurement variance on bias, and on how best to make the necessary trade-off between the  
6 loss of information (e.g. removing variables that exhibit significant cross-sample variance) and  
7 reducing bias in the results. In particular, one avenue for future research is to assess how the  
8 ratios of imputed data to used data within and between countries influence the imputation  
9 results.  
10  
11  
12  
13  
14  
15

16  
17 Even without simulation data though, our analyses suggest that it is important to some extent  
18 to evaluate the item distributions before imputation. The imputation example here was based  
19 on an assumption that the distributions were of the same shape. Naturally, in this case, this  
20 assumption was easily confirmed by looking at the real data (an option obviously unavailable  
21 in real situations). In cross-country surveys conducted with separate questionnaires, the  
22 planned missingness should thus be based on strong foreknowledge with a measurement  
23 model that has at least been verified in the research field of interest. Besides cross-country  
24 research, these same issues are essential if cross-survey designs are utilized for instance in  
25 cross-group consumer research. However, future simulation work should investigate in more  
26 depth the issue uncovered by our examination of the covariances amongst constructs across  
27 the different estimation methods. In particular, it is not clear yet whether accuracy of various  
28 imputation methods in replicating / retrieving the correct distribution shape of measurement  
29 items is a good indicator of whether inferences made of the relationships between latent  
30 constructs are likely to be correct. This should be a key area for future researchers.  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41

42  
43 Cross-survey designs may offer additional value for marketing scholars as they may help  
44 researchers in dealing with method biases. Podsakoff et al. (2003) suggest the use of  
45 measurement separation between related questions to minimize bias (item context effects,  
46 which refers to any influence on a question item due to its relation to the other items making up  
47 a survey (Wainer and Kiely, 1987; Tourangeau *et al.*, 2000), like item priming effects  
48 (Podsakoff *et al.*, 2003). Ostroff et al. (2002) show that the individual level method bias can  
49 result in aggregate level bias, and suggest also the use of a split-sample design to conduct  
50 cross-level and multilevel studies. Although many important issues remain for future research,  
51 to the best of our knowledge, this is the first article in the international marketing literature that  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 addresses the issues of cross-survey MI in the cross-country research context. As the use of  
4 cross-survey MI may be viewed as the norm rather than the exception in social science  
5 analysis (Rendall *et al.*, 2013) in the future, we hope to encourage international marketing  
6 scholars in considering such designs. However, the present paper shows there is much yet to  
7 be learned about how best to incorporate such designs and until conclusive simulation  
8 evidence is available, researchers should be cautions in using cross-survey MI designs where  
9 they are not assured of invariant measures (which may be a more common situation than  
10 many researchers assume).  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For Peer Review

## References

- Adams, L. M., and Darwin, G. (1982), "Solving the quandary between questionnaire length and response rate in educational research", *Research in Higher Education*, Vol. 17 No. 3, pp. 231-240.
- Baraldi, A. N. and Enders, C. K. (2010), "An introduction to modern missing data analyses", *Journal of School Psychology*, Vol. 48 No. 1, pp. 5-37.
- Berdie, D. R. (1989), "Reassessing the value of high response rates to mail surveys", *Marketing Research*, Vol. 1 No. 3, pp. 52-64.
- Brown, R. L. (1994), "Efficacy of the indirect approach for estimating structural equation models with missing data: A comparison of five methods", *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 1 No. 4, pp. 287-316.
- Chang, S. J., Van Witteloostuijn, A., and Eden, L. (2010), "From the editors: common method variance in international business research", *Journal of International Business Studies*, Vol. 41 No. 2, pp. 178-184.
- Covin, J. G., and Slevin, D. P. (1989), "Strategic management of small firms in hostile and benign environments", *Strategic Management Journal*, Vol. 10 No. 1, pp. 75-87.
- De Jong, M. G., Steenkamp, J-B.E.M., and Fox, J-P. (2007), "Relaxing measurement invariance in cross-national consumer research using a hierarchical IRT model", *Journal of Consumer Research*, Vol. 34, No. 2, pp. 260-278.
- Dillman, D. A., Sinclair, M. D., and Clark, J. R. (1993), "Effects of questionnaire length, respondent-friendly design, and a difficult question on response rates for occupant-addressed census mail surveys", *Public Opinion Quarterly*, Vol. 57 No. 3, pp. 289-304.
- Douglas, S. P. and Craig, C. S. (2006), "On improving the conceptual foundations of international marketing research", *Journal of International Marketing*, Vol. 14 No. 1, pp. 1-22.
- Enders, C.K. (2010), *Applied Missing Data Analysis*, Guilford Press, New York, NY.
- Fichman, M. and Cummings, J. N. (2003), "Multiple imputation for missing data: Making the most of what you know", *Organizational Research Methods*, Vol. 6 No. 3, pp. 282-308.
- Fornell, C. and Larcker, D. F. (1981), "Evaluating structural equation models with unobservable variables and measurement error", *Journal of Marketing Research*, Vol. 18 No. 1, pp. 39-50.
- Franke, G. R. and Richey, R. G. (2010), "Improving generalizations from multi-country comparisons in international business research", *Journal of International Business Studies*, Vol. 41 No. 8, pp. 1275-1293.
- Gelman, A., King, G., and Liu, C. (1998), "Not asked and not answered: Multiple imputation for multiple surveys", *Journal of the American Statistical Association*, Vol. 93 No.443, pp. 846-857.

1  
2  
3 Gibbons, J. D. and Chakraborti, S. (1992), *Nonparametric Statistical Inference*, Marcel Dekker, New  
4 York, NY.  
5

6  
7 Gottschall, A. C., West, S. G., and Enders, C. K. (2012), "A comparison of item-level and scale-level  
8 multiple imputation for questionnaire batteries", *Multivariate Behavioral Research*, Vol. 47 No. 1, pp. 1-  
9 25.  
10

11  
12 Graham, J. W., Taylor, B. J., Olchowski, A.E., and Cumsille, P. E. (2006), "Planned missing data  
13 designs in psychological research", *Psychological Methods*, Vol. 11 No. 4, pp. 323-343.  
14

15  
16 Hair, J. F., Anderson, R. E., Tatham, R. L., and Black, W. C. (1998), *Multivariate Data Analysis*,  
17 Prentice Hall, New Jersey.

18  
19 Hansen, J. D., Deitz, G. D., Tokman, M., Marino, L. D., and Weaver, K. M. (2011), "Cross-national  
20 invariance of the entrepreneurial orientation scale", *Journal of Business Venturing*, Vol. 26 No. 1, pp.  
21 61.  
22

23  
24 Hayduk, L. A. (1989), *Structural Equation Modeling with LISREL: Essentials and Advances*, The Johns  
25 Hopkins University Press, Baltimore.

26  
27 He, Y., Merz, M. A., and Alden, D. L. (2008), "Diffusion of measurement invariance assessment in  
28 cross-national empirical marketing research: perspectives from the literature and a survey of  
29 researchers", *Journal of International Marketing*, Vol. 16 No. 2, pp. 64-83.  
30

31  
32 Herzog, A. R., and Bachman, J. G. (1980), "Effects of questionnaire length on response quality", *Public  
33 Opinion Quarterly*, Vol. 45 No. 4, pp. 549-559.

34  
35 Honaker, J. and King, G. (2010), "What to do about missing values in time-series cross-section  
36 data", *American Journal of Political Science*, Vol. 54 No. 2, pp. 561-581.

37  
38 Horn, J.L. and McArdle, J.J. (1992), "A practical and theoretical guide to measurement invariance in  
39 aging research", *Experimental Aging Research*, Vol. 18 No. 3, pp. 117-144.  
40

41  
42 Hult, G.T.M., Ketchen, D. J., Griffith, D. A., Finnegan, C. A., Gonzalez-Padron, T., Harmancioglu, N.,  
43 Huang, Y., Talay, B. M., and Cavusgil, S. T. (2008), "Data equivalence in cross-cultural international  
44 business research: assessment and guidelines", *Journal of International Business Studies*, Vol. 39 No.  
45 6, pp. 1027-1044.

46  
47 Jambulingam, T., Kathuria, R., and Doucette, W. R. (2005), "Entrepreneurial orientation as a basis for  
48 classification within a service industry: the case of retail pharmacy industry", *Journal of Operations  
49 Management*, Vol. 23 No. 1, pp. 23-42.  
50

51  
52 Jaworski, B. J. and Kohli, A. K. (1993), "Market orientation: antecedents and consequences", *Journal of  
53 Marketing*, Vol. 57 No. 3, pp. 53-70.  
54

55  
56 Jöreskog, K. G. (1971), "Simultaneous factor analysis in several populations", *Psychometrika*, Vol. 36  
57 No. 4, pp. 409-426.  
58  
59  
60



- 1  
2  
3 Kelloway, E. K. (1998), *Using LISREL for Structural Equation Modeling*, Sage Publications, Thousand  
4 Oaks, CA.  
5  
6  
7 Knight, G. A. (1997), "Cross-cultural reliability and validity of a scale to measure firm entrepreneurial  
8 orientation", *Journal of Business Venturing*, Vol. 12 No. 3, pp. 213-25.  
9  
10 Kreiser, P. M., Marino, L. D., and Weaver, K. M. (2002), "Assessing the psychometric properties of the  
11 entrepreneurial orientation scale: A multi-country analysis", *Entrepreneurship Theory and Practice*, Vol.  
12 26 No. 4, pp. 71-94.  
13  
14 Little, R. J. (1988), "A test of missing completely at random for multivariate data with missing values",  
15 *Journal of the American Statistical Association*, Vol. 83 No. 404, pp. 1198-1202.  
16  
17 Little, R. J. A. and Rubin, D. B. (1987), *Statistical Analysis with Missing Data*, John Wiley, New York,  
18 NY.  
19  
20  
21 Littvay, L. (2009). "Questionnaire design considerations with planned missing data", *Review of*  
22 *Psychology*, Vol. 16 No. 2, pp. 103-113.  
23  
24 Malhotra, N. K., Agarwal, J., and Peterson, M. (1996), "Methodological issues in cross-cultural  
25 marketing research: a state-of-the-art review", *International Marketing Review*, Vol. 13 No. 5, pp. 7-43.  
26  
27 Mintu, A. T., Calantone, R. J., and Gassenheimer, J. B. (1994), "Towards improving cross-cultural  
28 research: Extending Churchill's research paradigm", *Journal of International Consumer Marketing*, Vol.  
29 7 No. 2, pp. 5-23.  
30  
31  
32 Narver, J. C. and Slater, S. F. (1990), "The effect of a market orientation on business profitability",  
33 *Journal of Marketing*, Vol. 54 No. 4, pp. 20-35.  
34  
35  
36 Olinsky, A., Chen, S., and Harlow, L. (2003), "The comparative efficacy of imputation methods for  
37 missing data in structural equation modeling", *European Journal of Operational Research*, Vol. 151 No.  
38 1, pp. 53-79.  
39  
40  
41 Oliveira, J. S., Cadogan, J. W., and Souchon, A. (2012), "Level of analysis in export performance  
42 research", *International Marketing Review*, Vol. 29 No. 1, pp. 114-127.  
43  
44 Olsson, U. (1979), "Maximum likelihood estimation of the polychoric correlation coefficient",  
45 *Psychometrika*, Vol. 44 No. 4, pp. 443-460.  
46  
47 Ostroff, C., Kinicki, A. J., and Clark, M. A. (2002), "Substantive and operational issues of response bias  
48 across levels of analysis: an example of climate-satisfaction relationships", *Journal of Applied*  
49 *Psychology*, Vol. 87 No. 2, pp. 355-368.  
50  
51  
52 Podsakoff, P. M., MacKenzie, S. B., Lee, J.Y., and Podsakoff, N. P. (2003), "Common method biases in  
53 behavioral research: A critical review of the literature and recommended remedies", *Journal of Applied*  
54 *Psychology*, Vol. 88 No. 5, pp. 879-903.  
55  
56 Reiter, J.P., Raghunathan, T. E., and Kinney, S. K. (2006), "The importance of modeling the sampling  
57 design in multiple imputation for missing data", *Survey Methodology*, Vol. 32 No. 2, pp. 143-149..  
58  
59  
60

- 1  
2  
3 Rendall, M. S., Ghosh-Dastidar, B., Weden, M. M., Baker, E. H., and Nazarov, Z. (2013), "Multiple  
4 imputation for combined-survey estimation with incomplete regressors in one but not both  
5 surveys", *Sociological Methods & Research*, Vol. 42 No. 4, pp. 483-530.  
6  
7  
8 Roszkowski, M. J., and Bean, A. G. (1990), "Believe it or not: Longer questionnaires have lower  
9 response rates", *Journal of Business and Psychology*, Vol. 4 No. 4, pp. 495-509.  
10  
11 Rubin, D. B. (1986), "Statistical matching using file concatenation with adjusted weights and multiple  
12 imputations", *Journal of Business & Economic Statistics*, Vol. 4 No. 1, pp. 87-94.  
13  
14 Rubin, D. B. (1987), *Multiple Imputation for Nonresponse in Surveys*, J. Wiley & Sons, New York, NY.  
15  
16 Rässler, S. (2002). Statistical matching. <http://tocs.ulb.tu-darmstadt.de/107415658.pdf>  
17  
18 Rässler, S., Koller, F., and Mäenpää, C. (2002), "A split questionnaire survey design applied to German  
19 media and consumer surveys", (No. 42b/2002). Diskussionspapiere//Friedrich-Alexander-Universität  
20 Erlangen-Nürnberg, Lehrstuhl für Statistik und Ökonometrie.  
21  
22  
23 Schafer, J. L. (1997), *Analysis of Incomplete Multivariate Data*, Chapman and Hall, New York, NY.  
24  
25 Schafer, J. L., and Graham, J.W. (2002), "Missing data: our view of the state of the art", *Psychological*  
26 *Methods*, Vol. 7 No. 2, pp. 147-177.  
27  
28 Schafer, J. L., and Olsen, M. K. (1998), "Multiple imputation for multivariate missing-data problems: A  
29 data analyst's perspective", *Multivariate Behavioral Research*, Vol. 33 No. 4, pp. 545-571.  
30  
31 Schenker, N., Raghunathan, T. E., and Bondarenko, I. (2010), "Improving on analyses of self-reported  
32 data in a large-scale health survey by using information from an examination-based survey",  
33 *Statistics in Medicine*, Vol. 29 No.5, pp. 533-545.  
34  
35  
36 Schunk, D. (2008), "A Markov chain Monte Carlo algorithm for multiple imputation in large surveys",  
37 *AStA Advances in Statistical Analysis*, Vol. 92 No. 1, pp. 101-114.  
38  
39 Shrive, F. M., Stuart, H., Quan, H. and Ghali, W. A. (2006), "Dealing with missing data in a multi-  
40 question depression scale: a comparison of imputation methods", *BMC Medical Research*  
41 *Methodology*, Vol. 6 No. 57.  
42  
43  
44 Steenkamp, J. B. E., and Baumgartner, H. (1998), "Assessing measurement invariance in cross-  
45 national consumer research", *Journal of Consumer Research*, Vol. 25 No. 1, pp. 78-107.  
46  
47  
48 Sundqvist, S., Kyläheiko, K., Kuivalainen, O., and Cadogan, J.W. (2012), "Kirznerian and  
49 Schumpeterian entrepreneurial-oriented behavior in turbulent export markets", *International Marketing*  
50 *Review*, Vol. 29 No. 2, pp. 203-219.  
51  
52  
53 Tighe, E., Livert, D., Barnett, M., and Saxe, L. (2010), "Cross-survey analysis to estimate low-incidence  
54 religious groups", *Sociological Methods & Research*, Vol. 39 No. 1, pp. 56-82.  
55  
56  
57 Tourangeau, R., Rips, L. J., and Rasinski, K. (2000), *The Psychology of Survey Response*, Cambridge  
58 University Press, New York, NY.  
59  
60



1  
2  
3 Tsikriktsis, N. (2005), "A review of techniques for treating missing data in OM survey research", *Journal*  
4 *of Operations Management*, Vol. 24 No. 1, pp. 53-62.

5  
6  
7 van Buuren, S. (2010), "Item imputation without specifying scale structure", *Methodology*, Vol. 6 No. 1,  
8 pp. 31-36.

9  
10 Verleye, G. (1997), "Missing at random data problems and maximum likelihood structural equation  
11 modeling", IPD Working Paper 1997-3. Vrije Universiteit Brussel. Centrum voor Statistiek en Wiskunde.

12  
13 Von Hippel, P. T. (2007), "Regression with missing Y's: An improved strategy for analyzing multiply  
14 imputed data", *Sociological Methodology*, Vol. 37 No. 1, pp. 83-117.

15  
16  
17 Wainer, H. and Kiely, G. L. (1987), "Item clusters and computerized adaptive testing: A case for  
18 testlets", *Journal of Educational Measurement*, Vol. 24 No. 3, pp. 185-201.

19  
20 Ziegelmeier, M. (2013), "Illuminate the unknown: evaluation of imputation procedures based on the  
21 SAVE survey", *AStA Advances in Statistical Analysis*, Vol. 97 No. 1, pp. 49-76.

22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Peer Review

## Item correlation comparison

		compag1	compag2	compag3	compag4	auton2	auton3	auton4	auton5
compag2	Original	0.474							
	A	0.288							
	B	0.317							
	C	-							
compag3	Original	0.223	0.244						
	A	0.139	0.202						
	B	0.079	0.232						
	C	-	-						
compag4	Original	0.513	0.563	0.301					
	A	0.362	0.515	0.327					
	B	0.357	0.496	0.298					
	C	-	0.510	-					
auton2	Original	-0.13	-0.128	-0.109	-0.105				
	A	0.016	0.032	-0.002	0.052				
	B	-0.014	0.004	-0.022	-0.008				
	C	-	-	-	-				
auton3	Original	-0.109	-0.060	-0.156	-0.038	0.585			
	A	0.001	-0.053	0.010	-0.047	0.345			
	B	-0.024	0.017	-0.005	-0.020	0.390			
	C	-	-	-	-	-			
auton4	Original	-0.077	-0.001	-0.206	-0.038	0.493	0.782		
	A	0.008	-0.045	0.029	-0.017	0.222	0.535		
	B	-0.049	-0.001	-0.009	-0.012	0.283	0.567		
	C	-	-0.060	-	-0.023	-	-		
auton5	Original	-0.026	-0.037	-0.124	-0.040	0.496	0.586	0.674	
	A	-0.018	-0.073	0.027	0.000	0.22	0.328	0.374	
	B	-0.011	0.030	-0.014	0.021	0.189	0.274	0.377	
	C	-	-0.057	-	-0.072	-	-	0.413	
proact1	Original	0.161	0.227	0.245	0.33	-0.143	-0.185	-0.194	-0.148
	A	0.212	0.268	0.165	0.37	-0.005	-0.062	-0.021	-0.045
	B	-	-	-	-	-	-	-	-
	C	-	-	-	-	-	-	-	-
proact2	Original	0.224	0.278	0.231	0.399	-0.211	-0.206	-0.199	-0.179
	A	0.135	0.232	0.105	0.359	0.003	-0.084	-0.112	-0.073
	B	0.187	0.234	0.140	0.376	-0.025	-0.101	-0.075	-0.036
	C	-	0.250	-	0.343	-	-	-0.098	-0.134
proact3	Original	0.288	0.312	0.212	0.464	-0.163	-0.213	-0.204	-0.161
	A	0.109	0.299	0.086	0.365	0.025	-0.089	-0.14	-0.064
	B	0.175	0.299	0.182	0.412	-0.019	-0.025	-0.046	0.029
	C	-	0.368	-	0.427	-	-	-0.108	-0.106

		compag1	compag2	compag3	compag4	auton2	auton3	auton4	auton5
inno1	Original	0.372	0.291	0.252	0.46	-0.117	-0.153	-0.185	-0.067
	A	0.118	<i>0.265</i>	0.082	0.332	0.071	-0.059	<i>-0.088</i>	<i>-0.057</i>
	B	<i>0.157</i>	<i>0.265</i>	<i>0.153</i>	<i>0.379</i>	0.063	-0.026	-0.060	-0.013
	C	-	<i>0.317</i>	-	0.356	-	-	-0.070	-0.135
inno2	Original	0.364	0.305	0.22	0.46	-0.139	-0.214	-0.217	-0.181
	A	0.137	0.243	0.08	0.324	0.09	-0.087	<i>-0.037</i>	-0.063
	B	0.146	0.242	<i>0.130</i>	<i>0.399</i>	0.049	-0.043	0.008	0.018
	C	-	<i>0.261</i>	-	0.353	-	-	-0.004	<i>-0.107</i>
inno3	Original	0.335	0.299	0.22	0.438	-0.128	-0.24	-0.213	-0.156
	A	0.15	<i>0.312</i>	0.079	0.349	0.068	-0.102	<i>-0.074</i>	-0.101
	B	0.138	0.281	0.134	<i>0.395</i>	0.027	-0.023	0.001	0.003
	C	-	0.321	-	0.372	-	-	-0.043	<i>-0.116</i>
risk_t1	Original	0.173	0.138	0.105	0.251	-0.166	-0.234	-0.144	-0.108
	A	0.171	0.258	-0.011	0.209	0.084	0.056	-0.053	-0.102
	B	-	-	-	-	-	-	-	-
	C	-	-	-	-	-	-	-	-
risk_t2	Original	0.159	0.148	0.13	0.275	-0.139	-0.205	-0.080	-0.103
	A	<i>0.145</i>	0.226	-0.029	0.181	0.13	0.040	0.010	-0.061
	B	0.098	0.232	-0.025	<i>0.205</i>	0.092	0.032	0.006	0.014
	C	-	<i>0.203</i>	-	0.179	-	-	0.026	<i>-0.091</i>
risk_t3	Original	0.19	0.154	0.112	0.282	-0.154	-0.216	-0.127	-0.164
	A	<i>0.113</i>	0.232	-0.034	0.169	0.103	0.011	<i>-0.029</i>	-0.067
	B	0.067	0.245	-0.008	<i>0.219</i>	0.077	0.010	0.005	-0.006
	C	-	<i>0.198</i>	-	0.170	-	-	-0.014	<i>-0.070</i>

\* The closest correlation coefficient is highlighted.