RESEARCH ARTICLE

Ensuring the validity of private forest owner typologies by controlling for response style bias and the robustness of statistical methods

Andrej Ficko* and Andrej Bončina

Department of Forestry and Renewable Forest Resources, University of Ljubljana, Biotechnical Faculty, Vecna pot 83, 1000 Ljubljana, Slovenia

(Received 2 April 2013; accepted 12 August 2013)

In survey-based segmentation of forest owners, two threats to the validity of results have largely been ignored: (1) response style bias and (2) the robustness of the statistical methods. This study demonstrates response style bias detection, presents an approach for correcting for acquiescence – the systematic tendency to agree with survey items, and explores the sensitivity of a probabilistic clustering algorithm to requirements for the validity of the typology. Structural equation modeling and Monte Carlo data generation techniques were employed to detect acquiescence and estimate its effect on construct validity. A survey of the relevance of management information for private forest owners (N = 364) was used as an example. Although acquiescence was confirmed, it had minor effect on the results and no effect on the substantive construct. Uncertainty about the number of forest owner types and membership can be reduced by using probabilistic clustering and observing the number of clusters while changing the requirements for the validity of clusters. The expectation maximization algorithm proved to be robust even to stringent requirements for the validity of clusters. By controlling for response style and the robustness of statistical methods, the validity of private forest owner typologies can be better ensured.

Keywords: acquiescence; Monte Carlo simulation; non-industrial private forest owners; response style bias; structural equation modeling; survey

1. Introduction

Surveys are one of the most frequently used instruments of measurement in social research in forestry. A researcher should consider several issues to ensure the validity of the results, for example, selection of the type of survey best suited to the problem domain and target population and design of the questionnaire to avoid biases in advance. Social scientists have raised some important concerns regarding possible biases in responses and their influences on construct validity, but these have been largely ignored by their counterparts in forestry despite the fact that both parties investigate a comparable population and use similar research design and data processing. To our knowledge, there are only a few studies on the social aspects of forestry that have recognized the potential threats to the validity of results due to bias in the input data or insufficient methodological rigor during the analysis (e.g. Egan & Jones 1993, 1995).

In social studies in forestry, private forest owner segmentation has long been popular for describing the diversity of private ownership. The number of published forest owner typologies increased after Kuuluvainen et al. (1996) pioneered quantitative methods of market segmentation in forestry. In quantitative segmentation, the analyst should account for two main uncertainties (Creswell 2003): (1) uncertainty about whether responses reflect the real opinion of a respondent or are biased (respondent uncertainty) and (2) uncertainty about whether the final segmentation of owners into the number of (usually disjoint) sets corresponds to reality, that is, model-reality consistency (Bollen 1989). Related issues include uncertainty about the number of customer segments, their meaning, and the fuzziness of membership (analyst uncertainty).

Private forest owner typologists have typically made the assumption that respondents know the answers to the questions and that their responses are an accurate reflection of their opinions. However, several behavioral, marketing, and sociological studies (cf. Weijters 2006; Van Vaerenberg & Thomas 2012) have found evidence of systematic response bias. Such consistent responding to items on the basis other than that the items were designed for has been referred to as response style (Paulhus 1991). Three common response styles have been identified (Paulhus 1991): the acquiescence response style (ARS), or the tendency to agree with the item irrespective of the content of that item; the disacquiescence response style (DARS), or consistent disagreement with the items irrespective of their content; and extreme
responding (ERS), which manifests as a preference for extreme response categories. Other common response styles include mid-point responding (MRS), or the tendency to use the middle response category and noncontingent responding (NCR), or responding that is careless, random, or non-purposeful (Van Vaerenbergh & Thomas 2012).

Biased responding may be linked to several external and internal stimuli (e.g. Baumgartner & Steenkamp 2001; Van Vaerenbergh & Thomas 2012). Inter alia, it may depend on an individual’s risk attitudes (Hofstede 2001); it may be influenced by social norms (e.g. the respondents may approve behavior that is socially desirable); or it may be related to the demographic variables and personality characteristics of a respondent. A lack of interest in the topic (“yeah answers”) may also lead to bias. In any case, failing to control for response style may lead to invalid research conclusions.

When a respondent recognizes his/her uncertainty, it can be quantified directly with a follow-up rating question on certainty immediately after the valuation question. Several approaches have been developed to account for self-reported uncertainty in contingent valuation (see e.g. Shaikh et al. 2007). Alternatively, Hujala et al. (2009) added the “I don’t know” category to the original Likert scale to control for respondent uncertainty and later eliminated these responses to account for self-reported uncertainty. However, such approaches still rely on a respondent’s self-reports and cannot diagnose the latent bias of a respondent.

To diagnose latent response style behavior, several techniques have been developed in behavioral, social, and marketing research (Van Vaerenbergh & Thomas 2012). For instance, methods based on response style indices (e.g. Bachman & O’Malley 1984; Reynolds & Smith 2010) are able to detect multiple types of response style and eliminate bias on an individual level, but fail to distinguish clearly between the response style and the content (Baumgartner & Steenkamp 2001; De Beuckelaer et al. 2010). In addition, the convergent validity of methods based on indices and more advanced methods for response style detection are not always secured (De Beuckelaer et al. 2010). Most response style diagnostics, however, are based on the assumption that if biased behavior exists, it can be identified as a common response style factor that equally loads on all items independent of their content (Billiet & McClendon 2000; Welkenhuyzen-Gybelts et al. 2003). This is the rationale used in our study and further described in the methods section.

Billiet and McClendon (2000) developed a procedure for the detection of and correction for acquiescence when modeling a construct. However, they did not present how to eliminate acquiescence from the raw data if any analysis other than construct modeling is required. Since our aim was to examine the response style effect on the identification of private forest owner segments, we further developed Billiet and McClendon’s procedure to eliminate ARS bias from the raw data.

The second source of uncertainty in the segmentation of forest owners is the analyst’s uncertainty about the model-reality consistency. In conventional approaches to forest owner segmentation (the Frequentist approach; Kangas & Kangas 2004), the analyst reports uncertainty with probability statements to convey scientific uncertainty after statistical modeling (e.g. with p-values). In the alternative approach (the Bayesian approach), the analyst reports the certainty with “a number between 0 and 1 that conveys the strength of belief or weight of evidence for some particular conjecture or hypothesis” (Ghazoul & McAllister 2003). The latter approach has several advantages in customer segmentation (e.g. fewer segments, cluster membership is determined with probabilities, multi-objectiveness is inherent to members of all groups; Magidson & Vermund 2002; Ficko & Boncina 2013b).

The aims of this research are: (1) to demonstrate Billiet and McCledon’s approach for the detection of response style bias in the field of forestry, (2) to develop a procedure for estimating the effect of response style bias in the event of response style contamination, (3) to explore the robustness of the probabilistic clustering algorithm to different requirements for the validity of private forest owner typology, and (4) to discuss the benefits of accounting for respondent and analyst uncertainty in private forest owner segmentation.

2. Methods

2.1. Sample survey design and preliminary analysis

We used responses from face-to-face interviews with 364 Slovenian private forest owners in the northern part of Slovenia (see Ficko & Boncina 2013b). Respondents were asked to rate the relevance of 19 items associated with management information for decision-making (Table 1, v1 to v19) using an equidistant five-point Likert scale (1 being not at all important, 5 being very important).

Like a marketer who uses the economic theory of market segmentation to maximize profit from selling a homogenous product to a market with heterogeneous demands (Wedel & Kamakura 1999), we attempted to identify major categories of information from the 19 items to structure the forest owners according to their information needs. Prior to this, we screened the distribution of response categories for each respondent and calculated various response style indices.
2.2. Theoretical framework for acquiescence response style detection

Billiet and McClendon (2000) developed the theoretical framework for detection of acquiescence in survey research based on the approaches of Mirowsky and Ross (1991) and Watson (1992). We followed their basic ideas, which can be summarized in four steps:

- If a substantial number of respondents systematically favors positive response categories irrespective of the content of the item, such behavior can be identified as a latent common factor referred to as the acquiescence response style (ARS) factor.
- When the set of items is semantically balanced (i.e. half of the items are positively worded, half of the items are negatively worded with respect to the construct being measured), the ARS factor can be identified directly as a factor that loads on all items with equal weight. When the set of items is not semantically balanced, but only maximally heterogeneous in content, the equivalence of such a factor to acquiescence can only be assumed. The heterogeneity of items is high if the average inter-item correlation is low. Baumgartner and Steenkamp (2001) reported an average inter-item correlation of 0.12, Johnson et al. (2005) reported 0.20, cit. in De Beuckelaer et al. (2010, p. 766). We therefore assumed that the responses might be contaminated by ARS.
- The identity of the ARS factor can be validated (in the case of a semantically balanced set of items) or confirmed (in the case of a maximally heterogeneous set of items) if it is found in two or more balanced sets of items measuring independent constructs, and the correlation between the ARS factor and the ARS indicator is high. The ARS indicator is the variable measuring the frequency of the very important and rather important response category selection.
- If ARS contaminates the responses, the model in which the ARS factor is incorporated should outperform the model consisting of content factors only in replicating the correlation matrix of the data, evidenced by better model fit.

2.3. Modeling acquiescence with structural equation modeling (SEM)

The existence of the ARS factor was tested with confirmatory factor analysis (CFA), which is a special type of structural equation modeling (SEM). Within the CFA we tested the hypothesis that the observed correlation matrix is equal to the correlation matrix implied by the hypothesized models (Models A and B, respectively, Figure 1). The measurement models consisted of a set of matrix equations (Bollen 1989, p. 17) representing relations between manifest ($v_i$) and latent variables ($\eta_i$ and $\delta_i$), with $\lambda_{vi}$ representing the loading of manifest variable $v_i$ to factor $\eta_i$ (Table 1):

$$v_i = \lambda_{vi}\eta_i + \delta_i \quad (1)$$

The models are presented modularly with path diagrams (Figure 1).

The content factors, their number, and hypothesized loadings of items on the factors were specified by the preliminary exploratory factor analysis since we had no theory to guide us in building the model. We specified six content factors and related them to those items that the exploratory factor analysis indicated and we thought the factors might load on (Model A, Figure 1; solid arrows only). The content factors were not allowed to correlate for theoretical reasons; they are intended to represent major, uncorrelated categories of information used in decision-making of different customer segments. Similarly, there was no theoretical reason to allow the correlation of residuals. In order to operate with a standardized scale, we set the scale of the factors using a constrained Fisher Scoring algorithm to produce a standardized solution. This algorithm standardized the variances of the factors (Hill & Lewicki 2007) and thus replaced the common practice of manual fixing of one path per factor to 1. Moreover, we analyzed correlations instead of covariances, resulting in a completely standardized path model and correctly calculated standard errors. All models were built and analyses conducted in the SEPATH module of STATISTICA 7.0 (Hill & Lewicki 2007).

To specify the model with the content factors and the ARS factor (Model B, Figure 1; solid and dotted arrows), we added a new factor, ARS, to model A and fixed all loadings of the items on this factor with the same value. By fixing the loadings of the items on ARS, we specified that all items are expected to be equally affected by the response style. The correlations between the six content factors and the ARS
factor were set to zero because there was no theoretical reason for correlation between the content and the style (e.g. Paulhus 1991).

To verify whether the ARS factor in Model B was indeed the ARS factor rather than an additional content factor, we added a new factor, “scoring for agreement” (N_agree1), to model B. We fixed the loading of the indicator variable measuring the frequency of the very important and rather important response category (vARS_index) on the factor N_agree1 to 1 and let N_agree1 correlate with the ARS factor and content factors. A negligible or insignificant correlation between N_agree1 and the content factors, but a strong correlation between N_agree1 and the
ARS factor would indicate that the ARS factor indeed measured acquiescence. We labeled the new model as model C (Figure 1; solid, dotted, and dashed arrows).

The $v_{ARS\text{index}}$ was constructed on two separate sets of items: the 19 items used for customer segmentation (set No. 1) and five items measuring the expectations of the extension services from the public forest service (set No. 2). If the ARS factor corresponds to the definition of stylistic responding, the correlation between the ARS factor and the two $v_{ARS\text{index}}$ constructed on two separate sets of items should be significant and stronger than the correlations between the ARS factor and the content factors.

To additionally verify the identity of the ARS factor, the $v_{ARS\text{index}}$ in model C was replaced with the disaquisition response style index ($v_{DARS\text{index}}$) and extreme response style index ($v_{ERS\text{index}}$), and the correlations between the ARS and scoring for disagreement factor and ARS and scoring for extreme response factor were estimated again.

### 2.4. Estimation procedure

Free parameters ($\lambda_{\eta_i \eta_j}$, $\psi_{\eta_i \eta_j}$, and $\delta_i$) were estimated with a discrepancy function, which is a summary measure of the size of the residuals in the model. When choosing the discrepancy function, we noted that the standard errors for parameter estimates as
well as the chi-square might be incorrect when using maximum likelihood estimation with non-normally distributed multivariate data (Savalei & Bentler 2006). In addition, we were also aware of the sensitivity of the chi-square statistic to sample size (e.g. Ullman 2006).

As an alternative to the robust parameter estimation procedures implemented in some of the structural equation modeling software packages (e.g. EQS, Bentler 2005), bootstrapping is an effective way for correcting the standard errors in SEM analysis (Bollen 1989; Ullman & Hancock 2001). Due to the indication of multivariate kurtosis in our data (normalized Mardia's (1970) coefficient > 3.00), we employed Monte Carlo bootstrapping to estimate the sampling distribution of model parameters and its standard errors as well as the distribution of the chi-square value. We used generalized least square estimation in the first five iterations, followed by maximum likelihood estimation until convergence (GLS-ML). We randomly drew a sample of size 364, 1000 times, with replacement, and each time fit the current model to the bootstrapped subsample.

Before deciding which discrepancy function to use, we compared the GLS-ML bootstrapping estimation with the asymptotic distribution-free estimation bootstrapping procedure (ADF), which is an alternative option in cases of multivariate non-normality (Savalei & Bentler 2006). The GLS-ML bootstrapping estimation resulted in a lower chi-square value than the ADF, which means that it was somewhat less restrictive to type I error, though the GLS-ML bootstrapped chi-square value was still higher than the critical value, where the hypothesis of perfect model fit would be accepted. More importantly, the GLS-ML bootstrapping estimation resulted in lower standard errors and smaller normalized residuals (max. ±3), making it a favorable estimation method for all our models. This empirical evidence supports the simulation studies that report better performance of ADF in large samples ( > 2500) or in rather simple models (e.g. Savalei & Bentler 2006; Ullman 2006). Neither of these two conditions was met in our case. Hence, all reported parameters in the models (Figures 2 and 3, and Table 2) are mean values obtained after GLS-ML bootstrapping 1000 times.

A theoretical perfect fit of the model to the data would result in a small chi-square value with a p-value of 1. The hypothesis of perfect fit was tested by comparing the GLS-ML bootstrapped chi-square at the corresponding df and p-value with the critical value at the corresponding df and p-value. In the goodness-of-fit quantification, we also considered model fit indices, which quantify how the pattern of correlations in the data is consistent with the specified model. Following the recommendations of Hu and Bentler (1999), we considered the Steiger-Lind Root Mean Squared Error of Approximation (RMSEA, Steiger 1990), the Goodness-of-Fit-Index (GFI) and Adjusted GFI (AGFI, Jöreskog & Sörbom 1993), the Comparative Fit Index (CFI, Bentler 1990), the Tucker and Lewis (1973) or Non-normed Fit Index (TLI), and the chi-square over the degrees of freedom ratio ($\chi^2$/df) (Bollen 1989). If the model fits perfectly, the GLS-ML bootstrapped chi-square value would be significantly lower than the critical value.

Table 2. Correlations ($\psi_{ij,hi}$) between the acquiescence response style factor (ARS), the content factors ($\eta_1$ to $\eta_6$), and the following factors: scoring for agreement (N_agree), scoring for disagreement (N_disagree), and scoring for extreme response (N_extreme), in two sets of items (No. 1 and No. 2). Correlations each time estimated with model C, other parameters of model C not shown.

|           | Set. No. 1 |               |      | Set. No. 2 |               |      |
|-----------|------------|---------------|------|------------|---------------|------|
|           | N_agree1   | N_disagree1   | N_extreme1 | N_agree2   | N_disagree2   | N_extreme2 |
| ARS       | 0.893*     | −0.893*       | 0.377*     | 0.305*     | −0.305*       | 0.130 |
| $\eta_1$  | 0.386*     | −0.386*       | 0.093      | 0.057      | −0.057        | 0.022 |
| $\eta_2$  | 0.181*     | −0.181*       | 0.280*     | −0.005     | 0.005         | 0.132 |
| $\eta_3$  | 0.106*     | −0.106*       | 0.055      | −0.146*    | 0.146*        | −0.159* |
| $\eta_4$  | 0.303*     | −0.303*       | −0.188*    | 0.260*     | −0.260*       | −0.229* |
| $\eta_5$  | 0.058*     | −0.058*       | 0.119*     | 0.004      | −0.004        | −0.025 |
| $\eta_6$  | 0.086*     | −0.086*       | 0.480*     | 0.113      | −0.113        | 0.222* |

Fit indices* for complete model C ($N = 364$):

|           |                |                |      | N_agree2 |                |                |
|-----------|----------------|----------------|------|----------|----------------|----------------|
| RMSEA     | 0.05           | 0.05           | 0.05 | 0.05     | 0.05           | 0.05           |
| GFI       | 0.93           | 0.93           | 0.93 | 0.93     | 0.93           | 0.93           |
| AGFI      | 0.90           | 0.90           | 0.90 | 0.91     | 0.91           | 0.90           |
| TLI       | 0.94           | 0.94           | 0.89 | 0.93     | 0.93           | 0.92           |
| CFI       | 0.95           | 0.95           | 0.93 | 0.94     | 0.94           | 0.94           |

*RMSEA (Steiger-Lind Root Mean Squared Error of Approximation, Steiger 1990); GFI, (Goodness-of-Fit-Index), AGFI (Adjusted Goodness-of-Fit-Index, both Jöreskog & Sörbom 1993); TLI (Tucker & Lewis 1973 or Non-normed Fit Index); CFI (Comparative Fit Index, Bentler 1990).

*Correlation significant at 0.05 level.
the indices should have a value of 1. Adequate model fit was evidenced by the indices having value of at least 0.90 (Hu & Bentler, 1999), except for the RMSEA, where a value of less than 0.06 (Hu & Bentler 1999), or 0.08 (McDonald & Ho 2002) is required. Statistically better model fit was determined by significant reductions in chi-square because model A and B are subsets of each other. Better model fit is also evidenced by higher values of fit indices and lower RMSEAs (Hu & Bentler 1999). All reported fit indices for models are mean values obtained after bootstrapping 1000 times.

2.5. Correcting for acquiescence

Once the ARS was detected, we proceeded with the following experiment to eliminate it from the raw data:

- Bearing in mind that the ARS inflates positive correlations and deflates negative ones between the items (Baumgartner & Steenkamp 2001; Van Vaerenbergh & Thomas 2012), we assumed that the observed positive correlations were more positive than they should be and the observed negative correlations were less negative or even positive.
- In confirmatory factor analysis, the model implied covariance matrix can be decomposed into matrices of factor loadings, factor covariances, and error covariances (Bollen 1989, p. 35, 236). In the standardized model with no correlations between the factors, the influence of factors on the correlation between two manifest variables reduces to the additive function of products of their loadings on those variables (Bollen 1989, see p. 192 for an illustration). This decomposition rule is fundamental for the next steps.
- If we managed to find a data-set whose correlation matrix perfectly fit the model A implied correlation matrix, then this data-set could be perfectly represented with exactly six content factors. Similarly, if we found a data-set whose correlation matrix perfectly replicated the model B implied correlation matrix, this data-set could be perfectly represented by six content factors and the ARS factor.
- To experimentally estimate the expected value of the ARS effect, we employed the Monte Carlo data generation technique. We simulated 1000 datasets from model A and 1000 datasets from model B, after A and B had been parameterized with the mean values obtained by the bootstrapping estimation procedure described in the section 2.4. From the 1000 datasets generated from the parameterized model A, we selected the one whose correlation matrix fit perfectly ($p > 0.99$) to the model (Table 3, the correlation matrix is shown as the lower triangular matrix). The same procedure was repeated for model B; the correlation matrix reproduced from the parameterized model B is shown in Table 3, upper triangular matrix.
- Since model A is nested within model B, that is, model A can be obtained by constraining ARS factor loadings in model B to zero for an increase of one degree of freedom, the ARS factor is uncorrelated to the content factors, and factor variances are fixed, the contribution of ARS to the correlation was estimated by comparing the correlation matrices implied by models A and B.
- Subtracting the lower triangular matrix from Table 3 from the upper triangular matrix from the same table provided an estimation of the effect of ARS on correlations (Net ARS).

Additional attempts were made with the Monte Carlo data generation procedure to simulate the population with corrected correlations among the items and desired distribution. We used Choleski factorization on the correlation matrix to convert independent normal random numbers to multivariate normal numbers with a desired correlation structure, and Vale and Maurelli’s (1983) technique to transform multivariate normal numbers into variates with desired non-normal distribution. The pseudocode for the described procedures is available online in the supplemental data.

2.6. Analyst uncertainty – the probabilistic approach

We explored the robustness of the expectation maximization (EM) clustering algorithm (Dempster et al. 1977) to decision-maker requirements for the validity of the model. In addition to the desired minimum and maximum number of clusters, an analyst can also specify the desired validity of the clustering solution. This can be done by specifying the allowable smallest percentage decrease in the evaluation function in cross-validating the solution, and by setting the precision of the minimum increase of the evaluation function. While the latter is of less practical interest
Table 3. Above diagonal; correlation matrix of the items (v1 to v19) perfectly fitting ($p > 0.99$) to Model B, selected from 1000 Monte Carlo replications. Below diagonal; correlation matrix of the items (v1 to v19) perfectly fitting ($p > 0.99$) to Model A, selected from 1000 Monte Carlo replications ($N = 364$).

|     | v1  | v2  | v3  | v4  | v5  | v6  | v7  | v8  | v9  | v10 | v11 | v12 | v13 | v14 | v15 | v16 | v17 | v18 | v19 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| v1  | 1.00| 0.55| 0.41| 0.20| 0.09| 0.12| 0.14| 0.29| 0.27| 0.15| 0.15| 0.13| 0.33| 0.14| 0.04| 0.11| 0.17| 0.12| 0.09|
| v2  | 0.53| 1.00| 0.33| 0.20| 0.07| 0.08| 0.16| 0.27| 0.24| 0.11| 0.15| 0.13| 0.28| 0.07| 0.07| 0.08| 0.10| 0.07| 0.04|
| v3  | 0.38| 0.31| 1.00| 0.20| 0.10| 0.11| 0.10| 0.11| 0.14| 0.11| 0.16| 0.14| 0.22| 0.06| 0.01| 0.10| 0.13| 0.12| 0.08|
| v4  | 0.18| 0.17| 0.16| 1.00| 0.10| 0.06| 0.10| 0.14| 0.10| 0.06| 0.07| 0.07| 0.16| 0.07| 0.17| 0.04| 0.10| 0.08| 0.13|
| v5  | -0.02| -0.04| 0.00| 0.00| 1.00| 0.53| 0.31| 0.11| 0.10| 0.20| 0.12| 0.11| 0.12| 0.20| 0.18| 0.19| 0.17| 0.23| 0.14|
| v6  | 0.03| -0.02| 0.01| -0.04| 0.51| 1.00| 0.51| 0.12| 0.08| 0.27| 0.16| 0.12| 0.11| 0.12| 0.11| 0.12| 0.09| 0.10| 0.13|
| v7  | 0.07| 0.10| 0.00| 0.00| 0.26| 0.49| 1.00| 0.28| 0.22| 0.16| 0.18| 0.18| 0.23| 0.17| 0.08| 0.14| 0.20| 0.24| 0.12|
| v8  | 0.25| 0.23| 0.00| 0.03| 0.01| 0.01| 0.22| 1.00| 0.58| 0.21| 0.18| 0.17| 0.41| 0.32| 0.17| 0.22| 0.28| 0.11| 0.08|
| v9  | 0.23| 0.21| 0.03| -0.01| -0.01| -0.02| 0.18| 0.55| 1.00| 0.24| 0.15| 0.15| 0.31| 0.24| 0.11| 0.19| 0.22| 0.09| 0.12|
| v10 | 0.06| 0.04| 0.00| -0.05| 0.16| 0.23| 0.13| 0.17| 0.19| 1.00| 0.03| 0.03| 0.28| 0.23| 0.27| 0.32| 0.19| 0.19| 0.14|
| v11 | 0.04| 0.05| 0.05| -0.04| 0.01| 0.06| 0.07| 0.07| 0.03| -0.08| 1.00| 0.81| 0.21| 0.30| 0.11| 0.15| 0.13| 0.06| 0.03|
| v12 | 0.02| 0.03| 0.03| -0.04| 0.00| 0.01| 0.08| 0.06| 0.04| -0.08| 0.81| 1.00| 0.18| 0.28| 0.11| 0.15| 0.11| 0.05| 0.06|
| v13 | 0.29| 0.25| 0.14| 0.08| 0.03| 0.00| 0.15| 0.38| 0.27| 0.23| 0.09| 0.06| 1.00| 0.33| 0.23| 0.31| 0.24| 0.13| 0.10|
| v14 | 0.05| -0.01| -0.06| -0.04| 0.15| 0.03| 0.08| 0.28| 0.18| 0.20| 0.22| 0.20| 0.29| 1.00| 0.45| 0.50| 0.37| 0.10| 0.07|
| v15 | -0.07| -0.04| -0.10| 0.07| 0.13| -0.01| -0.04| 0.10| 0.00| 0.24| 0.00| 0.00| 0.17| 0.43| 1.00| 0.56| 0.33| 0.11| 0.10|
| v16 | -0.01| -0.04| -0.02| -0.06| 0.14| -0.01| 0.03| 0.15| 0.08| 0.29| 0.03| 0.03| 0.24| 0.48| 0.56| 1.00| 0.39| 0.26| 0.14|
| v17 | 0.08| 0.02| 0.01| -0.01| 0.10| 0.03| 0.12| 0.22| 0.15| 0.15| 0.01| 0.00| 0.18| 0.33| 0.29| 0.36| 1.00| 0.32| 0.20|
| v18 | 0.02| -0.03| 0.02| -0.02| 0.16| 0.26| 0.16| -0.01| -0.03| 0.11| -0.07| -0.08| 0.01| -0.01| 0.01| 0.18| 0.25| 1.00| 0.56|
| v19 | -0.02| -0.06| -0.02| 0.03| 0.04| 0.02| 0.01| -0.04| 0.01| 0.04| -0.09| -0.07| -0.02| -0.04| 0.01| 0.07| 0.13| 0.55| 1.00|
for policy-makers, the desired validity of the clustering solution is useful for typology users.

We simulated decision-maker requirements on the validity of the probabilistic model by decreasing the smallest percentage decrease in the average log likelihood of cases for the next cluster solution in steps of 0.5% points, examining whether the more stringent validity requirements would result in more clusters. The simulation of less stringent requirements was meaningless because the minimum number of clusters (i.e. 2) had already been reached at the initially specified value of 1% decrease of log-likelihood (Ficko & Boncina 2013b).

3. Results

3.1. Acquiescence response style (ARS) detection

The confirmatory factor analysis (CFA) of information forest owners use in management decision-making confirmed that different types of information can be reduced into six major categories of information (Figure 2). However, the hypothesis of perfect fit had to be rejected ($\chi^2 = 490.00$, the number of degrees of freedom (df) = 137, $p < 0.05$): model A fit the data only marginally well (RMSEA = 0.08, GFI = 0.91, AGFI = 0.87, CFI = 0.90, TLI = 0.87, $\chi^2$/df = 3.6). The normalized residuals were in the approximate interval $[-1, 4]$.

After adding the ARS factor to model A, loadings of the content factors dropped, but remained of the same sign; the ARS factor loaded on the items with 0.330 (Figure 3) and the model fit improved ($\chi^2 = 404.45$, df = 136) but remained imperfect ($p < 0.05$). The difference in the $\chi^2$ statistics between model A and B amounted to 85.55 for 1 df, which is highly significant, $p < 0.001$. The better fit of Model B compared to Model A was also indicated when comparing the fit indices (RMSEA = 0.07, GFI = 0.93, AGFI = 0.90, CFI = 0.94, TLI = 0.93, $\chi^2$/df = 3.0). The normalized residuals were in the desirable interval $[-3, 3]$. We may conclude that the model with the ARS factor explains the data significantly better than the model with content factors only. The results thus demonstrated that respondents showed a tendency to agree with the survey items irrespective of their contents.

The parameters in Model C confirmed that the ARS factor indeed measured acquiescent responding and rejected speculation that the ARS factor is just an additional content dimension. The identity of the ARS factor was confirmed by the significant and strong correlation between the ARS factor

![Figure 2](image-url). The standardized solution of model A: content factor loadings and residuals, all significant at $p < 0.05$ (Fit indices: RMSEA = 0.08, GFI = 0.91, AGFI = 0.87, CFI = 0.90, TLI = 0.87, $\chi^2$/df = 3.6).
and the scoring for agreement factor ($\psi = 0.893$, $p < 0.05$ for set No. 1, $\psi = 0.305$, $p < 0.05$ for set No. 2, Table 2). The correlations between each scoring for agreement factor and the ARS factor were higher than the correlations between each scoring for agreement factor and the content factors (Table 2).

When the scoring for agreement factor in model C was replaced with the scoring for disagreement factor (N_disagree1 or N_disagree2), which loaded on the disaquiescence response style index ($v_{\text{DARS index}}$) with 1, the correlation between the N_disagree1 or N_disagree2 and the ARS was negative. Further indication of the identity of the ARS factor is given by the low correlation between ARS and N_extreme1 ($\psi = 0.377$, $p < 0.05$) and low and insignificant correlation between ARS and N_extreme2 ($\psi = 0.130$, $p = 0.243$). Moderate and significant correlations between N_extreme1 and content factors 6 and 2 ($\psi = 0.480$ for 6, and $\psi = 0.280$ for 2, $p < 0.05$) could be explained by the fact that the category “very important” was included in the calculation of both indices, $v_{\text{ERS index}}$ and $v_{\text{ARS index}}$.

### 3.2. Correcting for acquiescence

In the Monte Carlo experiment, we perfectly simulated the responses with the amount of ARS contamination as model B specified ($\chi^2 = 85.41$, df = 136, $p > 0.99$). The generated responses with no ARS contamination also perfectly fit the model ($\chi^2 = 82.79$, df = 137, $p > 0.99$). The calculations of correlations in Table 3 were exact to six decimal places, but only 2 are shown).

The average inflation of the correlations due to ARS was low (mean = 0.09, standard deviation = 0.03). We may conclude that if we observed the correlations between two arbitrary items at the level of approximately 0.09, there would actually be no correlation between these two items. Analogously, if we concluded that there was no correlation between two items, these two items would actually be weakly negatively correlated.

Acquiescence had no effect on substantive construct (Table 4). Correction for acquiescence resulted in a more clear identification of the major categories of information forest owners use in management decision-making. The loadings of content factors on...
items characterizing them (i.e. items with loadings greater than 0.50, in bold text in Table 4) slightly increased, whereas the loadings that were negligible for the interpretation of the factors decreased or even changed sign. The cumulative variance in the decision-making of private forest owners decreased from 64.1 to 63.3% when the responses were corrected for acquiescence.

Unfortunately, the Monte Carlo generation of the 364 responses with the desired corrected correlations between the 19 items was not accurate enough in 1000 attempts. Differences between the simulated data-implied correlation matrix and the corrected correlation matrix exceeded the average size of the ARS. Therefore, we stopped with the experiment in which clustering of generated cases was intended to resemble the clustering of forest owners.

3.3. Probabilistic clustering

EM clustering proved to be robust to the analyst’s requirements for validity. The EM algorithm continued to consolidate forest owners into two types even if the decrease in average log-likelihood of cases was required to be relatively small (0.5% or more). When the alternative cluster solution was required to be better than the existing one by less than 0.5%, the number of clusters increased to four (Table 5).

4. Discussion and conclusions

4.1. Methodological issues

Even though we further developed Billiet and McClendon’s procedure, the individual’s ranking of importance of information irrespective of his/her tendency to agree remained unsolved. Our procedure only accounted for the aggregate level of response style bias by correcting the correlations among the items measuring the content factors. This may be a deficiency when the individual-level scores are of interest, for instance in psychological studies. However, in forest owner segmentation the aggregate-level scores are of primary interest; the analyst typically wants to know which groups of forest owners will emerge from the sample data and what their meaning is, not how an individual from the sample responded.

In addition to our procedure for correcting for acquiescence, one could also follow the rationale of a number of scholars in the field of marketing research (e.g. Greenleaf 1992; Baumgartner & Steenkamp).
Table 5. The influence of desired validity of clusters on their number in expectation maximization clustering. The desired validity specified as the smallest percentage decrease in the evaluation function in cross-validating the solution.

| Desired validity of clustering solution (smallest percentage decrease) | Index of predictive validity (training negative log-likelihood) | Number of clusters |
|------------------------------------------------------------------------|-----------------------------------------------------------------|-------------------|
| 1.00000                                                                | 8.10357                                                         | 2                 |
| 0.51700                                                                | 8.10357                                                         | 2                 |
| 0.51610                                                                | 8.10315                                                         | 2                 |
| 0.51601                                                                | 8.10315                                                         | 2                 |
| 0.51600                                                                | 7.87817                                                         | 4                 |
| 0.50000                                                                | 7.87817                                                         | 4                 |
| 10^-8                                                                   | 7.87817                                                         | 4                 |

We would also like to note that the Monte Carlo experiment for correcting the correlations is valid for descriptive purposes only. The main threat is that the sampling error of the correlation estimates remains unknown and thus the corrected correlations cannot be used for further statistical modeling. If modeling is to be continued, new confirmatory factor analysis should be employed with all variables included in the model simultaneously and corrected correlations should not be used as the input.

Nevertheless, when the response style behavior is left undiagnosed and uncorrected, the influence of biased responding to segmentation results can be simulated by skewing the distribution of the responses (for ARS and DARS) or by recoding the responses (for ERS) and continuing with the procedures using distorted data (Ficko & Boncina 2013a). Any type of severely biased responding would result in significantly different cluster membership assignment. This simulation study found that if strong response style bias in the data-set actually existed, biased response-based clustering would only reduce the uncertainty about the true clusters by 21.9% to 37.6%, depending on response style. We believe that the simulation of response style effects is strong enough to illustrate some pitfalls that might be encountered in private forest owner segmentation.

4.2. Significance for decision makers
The ARS should be of particular concern when it changes the sign of correlations between the items. Since numerous statistical techniques used in private forest owner segmentation (e.g. PCA, regression and cluster analysis) are influenced by the magnitude of correlations directly or indirectly, it is reasonable to pay more attention to methodological rigor, otherwise...
conclusions directed toward policy-makers might be invalid. In our case, only loadings of content factors that were small changed sign from positive to negative or vice versa after the ARS was removed, having virtually no impact on the content of clustering variables.

In the justification of the possible reasons for acquiescence, we can only draw from this empirical study and general conclusions on respondent behavior from social and marketing studies. First, rating the relevance of information seems logical to generate an optimistic view of relevance by default; affirmative behavior may arise from the rationale that more information is beneficial when making decisions because information decreases uncertainty. Second, if the respondent is uncertain about how to respond, agreeing with an item may be less ambiguous than selecting a middle response category (Johnson et al. 2005; Smith 2004). Third, consistent agreement may also be a sign of politeness in face-to-face interviews or unwillingness to take on the cognitive load that the rating requires (Baumgartner & Steenkamp 2001). However, there is no agreement on the effect of these stimuli; ARS is reported to be less likely in face-to-face interviews than in other modes of data collection (Van Vaerenbergh & Thomas 2012).

Weijters (2006) investigated different sources of stylistic responding that range from survey instrument-based stimuli to personal characteristics. Unfortunately, none of these stimuli have been empirically proven to influence private forest owners. Despite this, some response styles besides ARS can be hypothesized to be more likely to occur among forest owners, for example socially desirable behavior (Steenkamp et al. 2010). When interviewed, forest owners could claim to be more multiobjective than they really are, trying to conform to the socially desirable concept of sustainable and multipurpose forest management. This could be the case for many typologies based on self-reported management objectives with a close-ended format. The authors of these typologies can verify their validity, test for the response styles, and make contributions to advanced social studies in forestry.

There is also space for analysts to improve accuracies. In addition to the pros and cons of the probabilistic clustering of private forest owners that have already been discussed (Ficko & Boncina 2013b), we would like to point to the added value that the simulation of desired validity of clusters can create. The desired validity of clusters can be specified in advance by the user of the typology. For instance, policy-makers can specify that the risk of an inaccurate clustering solution should be less than 5%. Alternatively, the analyst can investigate the validity by himself, as we have done. Since the number of clusters at the initial level of validity was already at the minimum, we only simulated more stringent validity requirements. Probabilistic clustering remained stable even under rather unrealistic decision-maker requirements for the validity of clusters, indicating that materialists and non-materialists from Ficko and Boncina (2013b) are valid groups despite minor contamination by ARS.

Our study pioneers response style detection and correction in private forest owner segmentation. However, it is based on only one data-set and controls for the effect of only one type of response style. We have no strong evidence that private forest owners are likely to respond with acquiescence or with any other style in general.

We may conclude that in addition to respondent uncertainty, which can be directly measured via self-reported uncertainty scores, detected and corrected with the aid of structural equation modeling when latent, or assessed by simulation, the validity of survey results can also be improved by examining the sensitivity of statistical methods employed during the analyses. The message to decision makers would then be more valid, and the private forest owner typologies would better serve as decision support systems for policy-makers.

Acknowledgments
We thank the Pahernik Foundation for financial support and Prof. Germà Coenders from the Faculty of Economics, University of Girona, Spain for providing assistance with the structural equation modeling.

Supplemental data
Supplemental data for this article can be accessed at http://dx.doi.org/10.1080/02827581.2013.837194.

References
Bachman JP, O’Malley PM. 1984. Yea-saying, nay-saying, and going to extremes: black-white differences in response styles. Public Opin Quart. 48:491–509.
Baumgartner H, Steenkamp J. 2001. Response styles in marketing research: a cross-national investigation. J Market Research. 38:143–156.
Bentler PM. 1990. Comparative fit indexes in structural models. Psych Bull. 107:238–246.
Bentler PM. 2005. EQS 6 structural equations program manual. Encino (CA): Multivariate Software.
Billiet JB, McClendon MJ. 2000. Modeling acquiescence in measurement models for two balanced sets of items. Struct Equ Model. 7:608–628.
Bollen KA. 1989. Structural equations with latent variables. New York: Wiley.
Creswell JW. 2003. Research design: qualitative, quantitative and mixed methods. Thousand Oaks (CA): Sage Publications.
