

## **A Permutation Test for Correlated Errors in Adjacent Questionnaire Items**

Laura A. Hildreth  
Iowa State University, Ames, IA 50011  
(515) 294-0473  
lahildre@iastate.edu

Ulrike Genschel  
Iowa State University, Ames IA 50011  
(515) 294-7766  
ulrike@iastate.edu

Frederick Lorenz  
Iowa State University, Ames, IA 50011  
(515) 294-7531  
folorenz@iastate.edu

Virginia Lesser  
Oregon State University, Corvallis, OR 97331  
(541) 737-3584  
lesser@science.oregonstate.edu

## **Abstract**

Response patterns are of importance to survey researchers because of the insight they provide into the thought processes respondents use to answer survey questions. In this paper we propose the use of structural equation modeling to examine response patterns and develop a permutation test to quantify the likelihood of observing a specific response pattern. Of interest is a response pattern where the response to the current item is conditioned on the respondent's answer to the immediately preceding item. This pattern manifests itself in the error structure of the survey items by resulting in larger correlations of the errors for adjacent items than for non-adjacent items. We illustrate the proposed method using data from the 2002 Oregon Survey of Roads and Highways and report SAS code which can be easily modified to examine other response patterns of interest.

# **A Permutation Test for Correlated Errors in Adjacent Questionnaire Items**

## ***1. Introduction***

Survey researchers are often interested in examining response patterns to provide insight into the processes respondents use to answer survey questions. Understanding these processes is important as they impact both the design and analysis of surveys. In this paper we propose the use of structural equation modeling (SEM) to examine response patterns and develop a permutation test for this purpose. Specifically, we use structural equation modeling to examine response patterns to “general-specific” questionnaires that contain both a series of items all pertaining to a related topic and a general question regarding the same topic. In general-specific questions the general question (G) either precedes (GS) or follows (SG) a series of specific items (S) that relate to the general question. To examine response patterns, the specific items are conceived of as manifestations of a single underlying latent factor, which is then regressed onto the latent factor associated with the general question where the general question is assumed to be measured without error. To illustrate our proposed method, i.e., using the SEM framework, we evaluate the hypothesis of a response pattern in which the response to a current item is conditioned on the respondent’s answer to the immediately preceding item. This response patterns manifests itself in the error structure of the specific items such that the correlations of the errors between adjacent items are larger than between non-adjacent items. To test this hypothesis, we use data from the 2002 Oregon Department of Transportation survey. We first estimate a baseline model that assumes uncorrelated errors in the measurement model and compare it to an alternative model that allows for the errors of adjacent items to be correlated. We then implement a permutation test to evaluate how likely the observed change in chi-square values from the null to the alternative model is under the sampling distribution of the change in

chi-square values from the null model to any alternative model. Note that the class of alternative models is restricted to models with the same latent variable model and equivalent degrees of freedom in the error structure of the measurement model. Although this sampling distribution is unknown, we can estimate it through simulation and implement a permutation test to evaluate the significance of the change in the model chi-square values.

The aim of this study is to extend the way in which simulation methods have previously been used in SEM. Simulation studies have proved invaluable in SEM research as many methodological advances were made possible by these studies (e.g. Curran, Bollen, Paxton, Kirby, and Chen, 2002; Fan and Fan, 2005; Fan and Sivo, 2007, 2009; Jackson, 2001). Such studies assess the consequences of model and data assumption violations and provide information about sample statistics whose distributions cannot be tracked analytically or have yet to be theoretically derived (Fan and Fan, 2005). Exemplar studies have evaluated the effects non-normality of latent variables and residual terms, estimation methods used to obtain structural parameter estimates, sample size, and model misspecifications on parameter estimates and model goodness of fit indices (e.g. Bentler and Satorra, 2010; Bentler and Yuan, 1999; Chen, 2007; Enders, 2001; Enders and Bandalos, 2001; Herzog and Boomsma, 2009; Lei and Lomax, 2005; Murphy, Beretvas, and Pituch, 2011; Olinsky, Chen, and Harlow, 2003; Ory and Mokhtarian, 2010). To the best of our knowledge, however, simulation studies have not been used to examine response patterns and thus the major contribution of this paper is to develop a permutation test and to present a SAS program for this purpose. Though we only examine one specific response pattern, both the methodology presented in this paper and the provided SAS program can be easily modified to examine other response patterns of interest to researchers.

## **2. Background: General Specific Questionnaires**

Questionnaires often contain a series of questions under a single preamble. Examples include “general-specific” or “part-whole” questions in which a general question (G) either precedes (GS) or follows (SG) a series of specific items (S) that relate to the general question. For example, psychologists have asked survey respondents to rate the overall quality of their life and the quality of specific domains of life such as work, family, friends and community (Campbell, Converse, and Rogers, 1972; McClendon and O’Brien, 1988). Similarly, researchers in community sociology, city planning, and marketing have asked respondents to rate their overall satisfaction with particular services such as local government services and facilities, and then also ask respondents to evaluate each specific service, such as streets, schools and police protection (DeMoranville and Bienstock, 2003; DeMoranville, Bienstock, and Judson, 2008; van de Walle and van Ryzin, 2011; Willits and Ke, 1995; Willits and Saltiel, 1995).

General-specific questionnaires have traditionally examined the effects of question order on responses to the general question by implementing “split-half” experiments in which respondents are randomly assigned one of two forms of the questionnaire. The resulting data are analyzed using either two sample t-tests or regression analysis. Under the two sample t-test approach, the means of the general question are compared for the two forms of the survey (GS or SG) where results from previous studies have indicated that the mean of the general question is more positive in the SG order than in the GS order (McClendon and O’Brien, 1987; Schul and Schiff, 1993; Willits and Ke, 1995). Under the regression approach, the specific items are conceived of as predictors, or “cause indicators” of the general question (Bollen and Davis, 2009; Bollen and Lennox, 1991; MacCallum and Browne, 1993) where an indicator variable for survey type (GS or SG) is used to model the order of the general and specific questions. The use of an indicator variable further allows the inclusion of interaction terms of the survey type and

specific items to the model when appropriate (Schul and Schiff, 1993). This approach enables researchers to compare estimated regression coefficients and  $R^2$  values within and between survey forms where previous studies have indicated higher  $R^2$  values for the SG order than for the GS order (McClendon and O'Brien, 1987; Schul and Schiff, 1993; Willits and Ke, 1995). These findings have been explained as follows: when the general question is asked first (GS), respondents scan their memory for salient information, and tend to recall negative events; in contrast, respondents answering the SG order interpret the general question as a request to summarize the specific items, both good and bad (Baumeister, Bratslavsky, Finkenauer, and Vohs, 2003; Krosnick and Presser, 2010; and Schul and Schiff, 1993).

Although these approaches have provided valuable insights, they have several shortcomings. A main issue, for example, is that under the regression approach collinearity among the specific items makes it difficult to draw inferences about the relative importance of specific items on responses to the general question. Perhaps more importantly though, previous research on surveys with general-specific questions has focused on the effects of the specific items on the general question, but it has not addressed relationships among the specific items, which would provide information regarding response patterns.

### **3. Theoretical Framework**

One approach to examine the relationships among specific items is structural equation modeling (SEM). Figure 1 illustrates this approach by depicting an SEM model in which the specific items are conceived of as manifestations of a single underlying exogenous latent factor,  $\xi$ , and the general question is measured (presumably without error) by the latent variable  $\eta$ . We assume the general question is measured without error as we have no measure of reliability

allowing us to estimate the error variance (Bollen, 1989). Finally, the latent variable  $\eta$  is regressed onto (or correlated with) the latent variable  $\xi$  associated with the specific items.

Under the SEM framework there are two submodels: the measurement model that describes the relationship between each latent variable and its indicators, and the latent variable model that describes the relationships among the latent variables themselves. We assume that all variables are deviated from their means, allowing us to express the equations associated with the measurement model, i.e., the relationship between each of the  $i = 1, 2, \dots, k$  specific items ( $Q_i$ ) and the latent variable,  $\xi$ , as well as the relationship between the general question and the latent variable  $\eta$  as simple linear regressions:

$$Q_i = \lambda_i \xi + \delta_i, \quad \delta_i \sim N(0, \sigma_i^2)$$

$$G = 1.0\eta + 0.$$

The response to each specific item  $Q_i$  is regressed against the latent variable  $\xi$  where  $\lambda_i$  denotes the regression coefficient and  $\delta_i$  is the residual (Bollen, 1989, 2002; Brown 2006). Similarly, the response to the general question,  $G$ , is regressed against the latent variable  $\eta$  where the regression coefficient is set equal to one and the error term is set equal to zero due to the default assumption that the general question is measured without error.

The responses to the general question, conceived of as manifestations of the latent variable  $\eta$ , are then regressed against the latent variable  $\xi$ , as shown by the directed arrow in Figure 1. The equation associated with the latent variable model is the regression of  $\eta$  on  $\xi$ , written as a simple linear model:

$$\eta = \gamma \xi + \zeta \quad \zeta \sim N(0, \sigma_\zeta^2)$$

where  $\gamma$  denotes the regression coefficient and  $\zeta$  denotes the latent error in equation such that  $\zeta \sim N(0, \sigma_\zeta^2)$ .

<FIGURE 1 ABOUT HERE>

Associated with the latent variable model are the covariance matrices  $\Psi$  and  $\Phi$ , where  $\Psi$  is the covariance matrix of errors in the latent variable equations and  $\Phi$  is the covariance matrix of the exogenous latent variables. In the context of Figure 1,  $\Psi$  is a 1 x 1 matrix of the error variance associated with latent error term  $\zeta$  and  $\Phi$  is also a 1 x 1 matrix of the variance of the latent variable  $\xi$ . Similarly there are two covariance matrices associated with the measurement model,  $\Theta_\delta$  and  $\Theta_\varepsilon$ , the covariance matrices associated with the measurement errors of the indicators associated with the exogenous and endogenous latent variables, respectively. In the context of Figure 1,  $\Theta_\varepsilon$  is a 1 x 1 matrix of the error variance associated with the general question, which we assume, is equal to 0.  $\Theta_\delta$  is a  $k \times k$  covariance matrix containing the error variances associated with the measurement errors of the specific items on the main diagonal and the error covariances between the measurement errors of the specific items on the off-diagonal.

One implication of the SEM model is that the error terms of the specific indicators are assumed to be uncorrelated with each other as the observed correlations among the specific items are assumed to be spurious due to the common latent variable (see visualization in Figure 1). The absence of arrows linking the error terms associated with the specific items implies that  $\text{cov}(\delta_i, \delta_j) = 0$  such that  $\Theta_\delta$  is a diagonal matrix of error variances:



$$\Theta_{\delta} = \begin{pmatrix} \theta_{11} & & & & & & \\ 0 & \theta_{22} & & & & & \\ 0 & 0 & \theta_{33} & & & & \\ 0 & 0 & 0 & \theta_{44} & & & \\ 0 & 0 & 0 & 0 & \theta_{55} & & \\ 0 & 0 & 0 & 0 & 0 & \theta_{66} & \end{pmatrix} .$$

In practice, however, it is possible that the latent variable does not absorb all correlations between error terms leading to the possibility that at least some of the off-diagonal elements of  $\Theta_{\delta}$  take on values other than zero in turn implying that there is a common source of variability between  $\delta_i$  and  $\delta_j$  beyond that due to the common latent variable  $\xi$ . In this study, we hypothesize and examine a specific pattern in the correlations among the error terms, which suggests that the unique variance in the response to one specific item is shaped by respondents' answers to the immediately preceding specific item. This results in correlated errors between the adjacent specific items, as denoted by the parameters  $\theta_{21}, \dots$ , and  $\theta_{k(k-1)}$  in Figure 2. Associated with the error terms of this model is the covariance matrix:

$$\Theta_{\delta} = \begin{pmatrix} \theta_{11} & & & & & & \\ \theta_{21} & \theta_{22} & & & & & \\ 0 & \theta_{32} & \theta_{33} & & & & \\ 0 & 0 & \theta_{43} & \theta_{44} & & & \\ 0 & 0 & 0 & \theta_{54} & \theta_{55} & & \\ 0 & 0 & 0 & 0 & \theta_{65} & \theta_{66} & \end{pmatrix} .$$

Because this may be only one of a number of possible response patterns we then further hypothesize that this response pattern is more common than any other response pattern implying that the correlations between error terms of adjacent items are significantly larger than between all other combinations of error terms associated with the specific items. We examine this hypothesis using a permutation test of the error structure of the specific items. Evidence in favor

of this hypothesis implies that respondents shape their response to the current item primarily on the most recent item as opposed to other previous items. This pattern might be more likely in situations where respondents have not developed strong opinions about the items. Under such conditions, respondents are more likely to take cues from the context in which questions are asked, and one context is the immediately preceding specific item.

<FIGURE 2 ABOUT HERE>

#### **4. Data and Methods**

The data used in our application come from the 2002 Oregon Survey of Roads and Highways. Oregon is divided into 16 transportation administration districts and, in each district, a random sample of Oregon drivers was selected and randomly mailed either a GS or SG form of the questionnaire. The general question asked “Overall, please rate how satisfied or dissatisfied you are with the conditions of those highways you have used in the past 12 months.” The responses were recoded on a scale from very satisfied (5) to very dissatisfied (1). The specific items, in order, asked about litter, rest areas, control of roadside vegetation, the gravel edges, guard rails and roadside guide posts. They were each coded on a scale from excellent (4) to poor (1). In 2002, of the 4100 mailed questionnaires 1746 were returned.

The two models associated with Figures 1 and 2, where  $k = 6$ , are of particular interest. The first model, which we refer to as the baseline or null model,  $M_0$ , assumes that all off-diagonal elements of the covariance matrix of the measurement errors are zero, implying that there is no systematic response pattern in the errors associated with the adjacent items. The second model, referred to as the model of adjacent correlations,  $M_1$ , eases this restriction and estimates the elements  $\theta_{21}, \theta_{32}, \dots, \theta_{65}$  and thus implying that the response to an item is conditioned on the previous item. The results of these two models, including estimated factor

loadings, estimated path coefficients, estimated correlations between adjacent errors for  $M_1$ , and model diagnostics such as the model chi-square values, the RMSEA (Root Mean Square Error of Approximation), and TLI (Tucker Lewis Index) are found in Table 1. The model diagnostics indicate that our models are appropriately specified and suggest that the adjacent correlations model fit the data better than the null model. When comparing the null model to the adjacent correlations model, the reduction in the model goodness of fit chi-square statistic is sufficiently large to conclude that the first order error model  $M_1$  fits the data significantly better than the null model. This is also reflected in the RMSEA value which decreases almost in half and the TLI value which increases by nearly 0.08.

For both the GS and SG versions of models  $M_0$  and  $M_1$ , the factor loadings are all above 0.5 and significant (see Table 1). The regression coefficient of the specific factor on the general factor is significant for all models reported and is smaller for the GS version than the SG version for both models. The estimated correlations between error terms are significant with the exception of  $\hat{\theta}_{32}$  and  $\hat{\theta}_{43}$  for the GS version, which indicates that they do not share a common source of variability. This suggests that, for whatever reason, respondents do not appear to have answered the 4<sup>th</sup> item (on gravel edges) in the list of specific items by taking cues from the 3<sup>rd</sup> item (on roadside vegetation). The two items may have been different enough to interrupt the response set.

<TABLE 1 ABOUT HERE>

Although the chi-square value decreases when allowing for first order correlated errors, the chi-square value would decrease if any set of correlated errors were freely estimated. This leads to the following question: Is the change in the estimated chi-square from the null model to the first order model larger than the reduction in the chi-square from the null model to any other

model with the same number of estimated correlated pairs of errors? To answer this question we adapt and implement the permutation test described below.

### **Permutation test procedure**

Because the sampling distribution of the reduction in chi-square values for our data is unknown, it must be estimated. This estimation is possible in the framework of a permutation test of the error structure. To conduct such a test we first define a null hypothesis as follows:  $H_0$ : the decrease in the model chi-square value when allowing for first order correlated error terms is not significantly larger than the decrease in the model chi-square value for any other model with the same number of correlated error terms. This hypothesis implies that respondents do not take more cues from the immediately preceding question to answer the current question than they take from any other question. If the null hypothesis is true, then any set of randomly chosen correlations of error terms could be added to the model and the reduction in chi-square associated with this model compared to the null model is not considerably different than the values shown in Table 1. To test the null hypothesis, the following steps, as adapted from Efron and Tibshirani (1998), are used:

*Step 1:* Decide on a test statistic and compute its value. Here, the test statistic of interest is the change (drop) in the model chi-square value from the null model to the model with correlated first order errors as found in Table 1.

*Step 2:* Construct the sampling distribution of the test statistic in step 1 as if the null hypothesis is true. For this step the data, the pairs of correlated errors are permuted under the assumption that the null hypothesis is true. In the present application, there exist six error terms associated with the specific questions leading to 15 unique pairs of correlated errors and 3003 distinct ways to randomly select 5 correlations from among 15 possible correlations without

replacement<sup>1</sup>. Using a SAS macro, described in Section 5, a set of five pairs of correlated errors is randomly chosen without replacement and is then added to the null model (five pairs were chosen as there are five pairs of first order correlated errors and thus the degrees of freedom of these models will be the same). Next, the reduction in the chi-square value associated with this model from the null model is calculated. This process is repeated a total of 999 times to approximate the sampling distribution of the reduction in chi-square value (999 was chosen to yield a total 1000 models including the model used in step 1).

*Step 3:* Count the number of cases that produce a test statistic at least as unusual as the observed test statistic in step 1. For this step the number of cases that had a reduction in chi-square as large as or larger than what was obtained in step 1 are counted. Because the change in the chi-square value for the model in step 1 is known, one is added to this number.

*Step 4:* Calculate the p-value by dividing the number found in step 3 by the total number of fitted models. As with step 3, one is added to the number of models run to account for the model run in the first step, leading to a total number of models equal to 1000.

The results for the permutation test are displayed in Figure 3. For each form, the reduction in chi-square for each fitted model is plotted where the vertical line represents the reduction in chi-square from the null model to the adjacent correlations model. The p-value is the area to the right of this line. For the GS questionnaire the p-value is 0.084 while for the SG questionnaire the p-value is 0.035. These results indicate that for the GS case there is weak evidence to suggest a systematic first order structure in the error terms while for the SG case there is stronger evidence of a systematic first order structure. Consequently, the statistical

---

<sup>1</sup> While it would potentially be computationally feasible in this example to compute all 3003 models, this need not be the case, especially when there are more than six indicators associated with a latent variable. For example, suppose there are 10 indicators, implying nine first order correlations and 45 possible pairs of correlated errors. This would lead to 886163135 different models, which is not computationally feasible.

pattern found in the data is consistent with a pattern respondents would produce if the response an item is conditioned by their response to the immediately preceding item but not any of the other items. In this case, the image we have is that survey respondents work through questionnaires in an almost linear fashion, answering each question with whatever information is available at that moment, and a substantial part of that information is their recall of the immediately preceding item and their response to it.

<FIGURE 3 ABOUT HERE>

## **5. SAS Code**

The previous analysis of Section 4 was conducted in SAS version 9.22 using PROC CALIS to estimate the models and a SAS macro (presented below) to conduct the permutation test. SAS was used instead of the more familiar LISREL, Mplus, SPSS (AMOS) or other SEM programs because it has the programming flexibility necessary to conduct the permutation test. While these other programs are capable of simulating data, they do not allow for simulating different model structures as done in our study. The SAS program used for our analysis is presented in Appendix A.

There are three major components to this SAS program. In the first part, a data set is created that specifies all possible pairs of correlated errors. The second part consists of a macro that randomly selects five of these pairs created in the first part and adds these to the null model with no correlations. The chi-square value associated with the executed model is saved. In the final part the simulated p-value is calculated using the chi-squared values from the simulated models.

### **Variable Names and Descriptions**

Before presenting the SAS code for this simulation, the variables and a brief description of the variables used in this analysis are given in Table 2. In relation to the conceptual model in Figure 1,  $q_{2a\_1}$  through  $q_{2f\_1}$  are represented by Q1 through Q6 in Figure 1, respectively while  $q_3$  is represented by the variable  $\eta_2$ .

<TABLE 2 ABOUT HERE>

**Component 1: Create Data Set of Possible Pairs of Correlated Errors.** A data set named `errors` is created first by specifying all possible pairs of correlated errors. As mentioned in Section 4 there are 15 unique correlations based on two distinct items from the list of Q1 through Q6 leading to the 15 observations in this data set.

**Component 2: SAS Macro.** The next step of the analysis entails randomly selecting five pairs of correlated errors without replacement, which we implemented through a macro. This macro consists of four distinct parts—randomly selecting five pairs of correlated errors from the data set `errors`, defining these pairs as macro variables, adding these pairs of correlated errors to the null model, and finally saving the chi-square value associated with the new model. To ensure that the same randomly selected errors are chosen for each year and survey version, we set the seed to be equal to the value of the associated simulated model (i.e. for the third model the seed is set to three) plus the constant 900 (chosen randomly). This ensures that the simulated p-values remain the same each time the program is run. To determine if this seed possibly led to anomalous results different constants besides 900 were randomly chosen and tested. The results of these simulated p-values are markedly similar to the aforementioned simulated p-values.

*Part 1: Choose Five Pairs of Correlated Errors.* Using PROC SURVEYSELECT, five observations are drawn from the data set `errors` using simple random sampling without replacement. These observations are then outputted to the data set `rs`.

*Part 2: Define the Randomly Chosen Pairs of Correlated Errors as Macro Variables.*

Because the randomly chosen pairs of correlated errors change for each simulation, they must be defined as macro variables. This is accomplished using two steps. In the first step, the data set `rs` is read into PROC IML in order to reformat the data set in a way that simplifies the defining of macro variables in the second step through the use of PROC SQL.

*Parts 3 and 4: Add Pairs of Correlated Errors to the Null Model and Save the Associated Chi-Square Value.* In this part of the macro, PROC CALIS is called to run the null model as specified by the user. The `pcov` statement is called and the randomly selected pairs of correlated errors are added to the null model using the macro variables defined in the previous part. The fit indices associated with this model are saved into the data set `semout2002`. Using PROC TRANSPOSE, the chi-square value for each simulation is saved in the data set `chisq2002`.

**Component 3: Calculate the p-value.** The last stage of this study is to calculate the simulated p-value through two steps. In the first step the null model and the adjacent correlations models are run. The associated chi-square values are saved using PROC TRANSPOSE in a process similar to the one used for the chi-square values in the macro. Secondly, these data sets and the data set `chisq2002` are read into PROC IML to calculate the p-value. Using PROC IML the change in chi-square from the null model to the adjacent correlations model is calculated as are the reduction in chi-square from the null model to each of the simulated models. The simulated p-value is then calculated as the proportion of changes that are as great as or greater than the drop from the null to the first order error model.

## **6. Discussion and Conclusions**



Understanding response patterns in surveys is a topic of interest to survey researchers as it impacts both the design and analysis of surveys. In this paper we use structural equation modeling to examine response patterns through the implementation of a permutation test of the error structure of the survey items in general-specific questionnaires. Though simulation studies have been used to examine many topics of interest in SEM, to the best of our knowledge they have not been used to examine specific response patterns in social surveys. In our particular application, we examine the presence of a first order response pattern in which respondent's answer to a given specific item is conditioned on their answer to the immediately preceding item. Through this analysis, we make two major contributions to the literature on surveys and SEM. First, is the development and implementation of a permutation test allowing us to test a research hypothesis regarding response patterns that, to our knowledge, has not previously been examined in the literature. Second, we provide the SAS program necessary to conduct such a test. This SAS program can be easily modified to examine other response patterns of interest.

We understand that the presented results have some limitations. First, the hypothesized pattern in the residuals is not the only possible pattern; for example, we could hypothesize that the error terms for the 2<sup>nd</sup> through last specific items are conditioned by the 1<sup>st</sup> item rather than by the immediately adjacent item. This hypothesis too can be tested with the presented SAS algorithm. Second, we acknowledge that this pattern may not be present in all surveys and also may depend upon the order of the specific items. We suspect that this type of response pattern is most likely to appear in surveys in which respondents do not have strong feelings about the survey topic. Under this scenario, they may take cues from other sources, including contextual elements of the questionnaire, and we hypothesized that the most proximal contextual factor was the immediately preceding specific item.

Though this paper gives results for only one survey during one year, we employed the procedure described in this paper with multiple years of data from the Oregon road quality survey, as well as with data regarding governmental services in rural Iowa towns. It is perhaps noteworthy that for the Oregon road quality survey the order of the specific items varied across survey years, thus allowing for different correlated errors among specific items. Regardless of the order of the specific items, evidence was found to suggest the presence of the response pattern of interest in the majority of these surveys indicating that the ordering of the specific items may not be able to mitigate the response pattern of respondents. All of these factors strengthen the likelihood that our findings are not idiosyncratic and the permutation test we develop and implement using the provided SAS program provides a useful way to examine this and other patterns. Identifying such patterns is crucial to better understand systematic sampling error that may create bias in response patterns and threaten the validity of sample surveys.

## References

- Baumeister, R. F., Bratslavsky, E., Finkenauer, C. and Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5, pp. 323 – 370.
- Bentler, P. M. and Satorra, A. (2010). Testing model nesting and equivalence. *Psychological Methods*, 15(2), pp. 111-123.
- Bentler, P. M. and Yuan, K. (1999). Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research*, 34(2), pp. 181-197.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, 53, pp. 605-634.
- Bollen, K. A. and Davis, W. R. (2009). Causal indicator models: Identification, estimation, and testing. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(3), pp. 498-522.
- Bollen, K. A. and Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110(2), pp. 305-314.
- Brown, Timothy A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.
- Campbell, A., Converse, P. E., and Rogers, W. L. (1976). *The quality of life in America: Perceptions, evaluations, and satisfactions*. New York: Russell Sage Foundation.
- Chen, F. F. (2007). The sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), pp. 464-504.
- Curran, P. J., Bollen, K. A., Paxton, P. Kirby, J. and Chen, F. (2002). The noncentral chi-square distribution in misspecified structural equation models: Finite sample results from a Monte Carlo simulation. *Multivariate Behavioral Research*, 37(1), pp. 1-36.
- DeMoranville, C. W. and Bienstock, C. C. (2003). Question order effects in measuring service quality. *International Journal of Research in Marketing*, 20, pp. 217-231.
- DeMoranville, C. W., Bienstock, C. C., and Judson, K. (2008). Using question order for predictive quality measures. *Journal of Services Marketing*, 22(3), pp. 253-262.
- Efron, B. and Tibshirani, R. J. (1998). *An introduction to the bootstrap*. USA: Chapman & Hall/CRC.
- Enders, C. K. (2001). The impact of nonnormality on full information maximum-likelihood

- estimation for structural equation models with missing data. *Psychological Methods*, 6(4), pp. 352-370.
- Enders, C. K. and Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 8(3), pp. 430-457.
- Fan, X. and Fan, X. (2005). Using SAS for Monte Carlo simulation research in SEM. *Structural Equation Modeling: A Multidisciplinary Journal*, 12(2), pp. 299-333.
- Fan, X. and Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, 42(3), pp. 509-529.
- Fan, X. and Sivo, S. A. (2009). Using goodness-of-fit indexes in assessing mean structure invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(1), pp. 54-69.
- Herzog, W. and Boomsma, A. (2009). Small-sample robust estimators of noncentrality-based and incremental model fit. *Structural Equation Modeling: A Multidisciplinary Journal*, 16(1), pp. 1-27.
- Jackson, D. L. (2001). Sample size and number of parameter estimates in maximum likelihood confirmatory factor analysis: A Monte Carlo investigation *Structural Equation Modeling: A Multidisciplinary Journal*, 8(2), pp. 205-223.
- Krosnick, J. A. and Presser, S. (2010). Question and questionnaire design. In P. V. Marsden and J. D. Wright (Eds.), *Handbook on Survey Research* (2nd ed., pp. 263—314). UK: Emerald Group Publishing Limited
- Lei, M. and Lomax, R. G. (2005). The effect of varying degrees of nonnormality in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, 12(1), pp. 1-27.
- MacCallum, R. C. and Browne, M. W. (1993). The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin*, 114(3), pp. 533-541.
- McClendon, K. J. and O'Brien, D. J. (1988). Question-order effects on the determinants of subjective well-being. *Public Opinion Quarterly*. 57, 351 – 364.
- Murphy, D. L., Beretvas, S. N., and Pituch, K. A. (2011). The effects of autocorrelation on the curve-of-factors growth model. *Structural Equation Modeling: A Multidisciplinary Journal*, 18(3), pp. 430-448.
- Olinsky, A. Chen, S. and Harlow, L. (2003). The comparative efficacy of imputation methods for missing data in structural equation modeling. *European Journal of Operational Research*, 151, pp. 53-79.

- Ory, D. T. and Mokhtarian, P. L. (2010). The impact of non-normality, sample size and estimation technique on goodness-of fit measures in structural equation modeling: Evidence from ten empirical models of travel behavior. *Quality and Quantity*, 44, pp. 427-445.
- Schul, Yaacov and Schiff, Miriam. 1993. Measuring satisfaction with organizations: Predictions from information accessibility. *Public Opinion Quarterly*, 57, pp. 536 – 551.
- van de Walle, S. and van Ryzin, G. G. (2011). The order of questions in a survey on citizen satisfaction with public services: Lessons from a split-ballot experiment. *Public Administration*, 89(4), pp. 1436-1450.
- Willits, F. K. and Ke, B. 1995. Part-whole question order effects: Views of rurality. *Public Opinion Quarterly*, 29, pp. 392 – 403.
- Willits, F. K. and Saltiel, J. (1995). Question order effects on subjective measures of quality of life: A two-state analysis. *Rural Sociology*, 60(4), pp. 654-665.

## Appendix A: SAS Code

```
/*Component 1: Construct data set of all unique pairs of correlated errors*/
data errors;
input q1 $ q2 $;
datalines;
q2a_1 q2b_1
q2a_1 q2c_1
q2a_1 q2d_1
q2a_1 q2e_1
q2a_1 q2f_1
q2b_1 q2c_1
q2b_1 q2d_1
q2b_1 q2e_1
q2b_1 q2f_1
q2c_1 q2d_1
q2c_1 q2e_1
q2c_1 q2f_1
q2d_1 q2e_1
q2d_1 q2f_1
q2e_1 q2f_1
;
run;

/*Component 2: SAS macro */
/*Define the macro*/
%MACRO Do_tcalis2002 (
datst, /*input data set*/
nrun /*number of simulations*/
);
options nonotes nosource nosource2 errors=0;
%DO I = 1 %TO &nrun;
%let seed=%eval(&I+900);

/*Part 1: Randomly select five pairs of correlated errors without
replacement*/
proc surveyselect data=errors method=SRS rep=&nrun sampsize=5 seed=&seed
out=rs noprint;
id _all_;
run;

/*Part 2: Define these pairs as macro variables*/
proc iml;
use rs;
read all var{'q1' 'q2'} into XM;
XMnew = XM[1,] || XM[2,] || XM[3,] || XM[4,] || XM[5,];
cname = {"qe1" "qe2" "qe3" "qe4" "qe5" "qe6" "qe7" "qe8" "qe9" "qe10"};
create store from XMnew [colname = cname];
append from XMnew;
quit;
proc sql noprint;
select *
into: qe1, :qe2, :qe3, :qe4, :qe5, :qe6, :qe7, :qe8, :qe9, :qe10
from store;
%put &qe1; %put &qe2; %put &qe3; %put &qe4; %put &qe5;
%put &qe6; %put &qe7; %put &qe8; %put &qe9; %put &qe10;
```

```

/*Parts 3 and 4: Add correlated errors to the null model and save the chi-
square value associated with this model*/
proc calis data = &datst cov outfit=semout2002 nop;
path
/*measurement model*/
q2a_1 <- specific = 1.0,
q2b_1 <- specific = lambda2,
q2c_1 <- specific = lambda3,
q2d_1 <- specific = lambda4,
q2e_1 <- specific = lambda5,
q2f_1 <- specific = lambda6,
/*structural model*/
specific -> q3 = b21;
/*randomly chosen pairs*/
pcov
&qe1 &qe2 = corr1,
&qe3 &qe4 = corr2,
&qe5 &qe6 = corr3,
&qe7 &qe8 = corr4,
&qe9 &qe10 = corr5;
by version;
run;
data semout2002 (keep=FitValue);
set semout2002;
where IndexCode = 203;
run;
proc transpose data=semout2002 out=newfit2002(drop=_LABEL_ _NAME_)
prefix=chisqv let;
run;
proc append base=chisq2002 data=newfit2002 FORCE;
run;
%END;
%MEND Do_tcalis2002;
%Do_tcalis2002(odot2002,999)

/*Component 3: Calculate simulated p-value*/
/*Run the null model with no constraints and save the chi-square value*/
proc calis data = odot2002 cov outfit=m0;
path
/*measurement model*/
q2a_1 <- specific = 1.0,
q2b_1 <- specific = lambda2,
q2c_1 <- specific = lambda3,
q2d_1 <- specific = lambda4,
q2e_1 <- specific = lambda5,
q2f_1 <- specific = lambda6,
/*structural model*/
specific -> q3 = b21;
by version;
run;
/*Run the model with added correlated first order errors and save the chi-
square value*/
proc calis data = odot2002 cov outfit=m1;
path
/*measurement model*/
q2a_1 <- specific = 1.0,

```

```

    q2b_1 <- specific = lambda2,
    q2c_1 <- specific = lambda3,
    q2d_1 <- specific = lambda4,
    q2e_1 <- specific = lambda5,
    q2f_1 <- specific = lambda6,
    /*structural model*/
    specific -> q3 = b21;
    pcov
    q2a_1 q2b_1 = corr1,
    q2b_1 q2c_1 = corr2,
    q2c_1 q2d_1 = corr3,
    q2d_1 q2e_1 = corr4,
    q2e_1 q2f_1 = corr5;
by version;
run;
/*keep only needed values in the output for M0 and M1*/
data m0 (keep=FitValue);
set m0;
where IndexCode = 203;
run;
proc transpose data=m0 out=m0t(drop=_LABEL_ _NAME_) prefix=chisqM0v let;
run;
data m1 (keep=FitValue);
set m1;
where IndexCode = 203;
run;
proc transpose data=m1 out=m1t(drop=_LABEL_ _NAME_) prefix=chisqM1v let;
run;
/*use PROC IML to calculate p-values*/
proc iml;
use m0t;
read all var _all_ into M0;
use m1t;
read all var _all_ into M1;
use chisq2002;
read all var _all_ into chisq;
changeM0M1 = M0 - M1;
changeM0M1v1 = changeM0M1[1,1];
changeM0M1v2 = changeM0M1[1,2];
change = M0 - chisq;
n = NROW(chisq);
v1p = J(n,1,0);
v2p = J(n,1,0);
do i = 1 to n;
if change[i,1] >= changeM0M1v1 then v1p[i,1] = 1;
if change[i,2] >= changeM0M1v2 then v2p[i,1] = 1;
end;
p1 = sum(v1p)/(n+1);
p2 = sum(v2p)/(n+1);
print p1 p2;
run;

```