BIOMETRICS 60, 241–248
March 2004

CONSULTANT'S FORUM

Testing for Marginal Independence between Two Categorical Variables with Multiple Responses

Christopher R. Bilder

Department of Statistics, University of Nebraska-Lincoln, Lincoln, Nebraska 68583, U.S.A.
email: cbilder3@unl.edu, http://www.chrisbilder.com

and

Thomas M. Loughin

Department of Statistics, Kansas State University, Manhattan, Kansas 66506, U.S.A.

SUMMARY. Questions that ask respondents to "choose all that apply" from a set of items occur frequently in surveys. Categorical variables that summarize this type of survey data are called both pick any/c variables and multiple-response categorical variables. It is often of interest to test for independence between two categorical variables. When both categorical variables can have multiple responses, traditional Pearson chi-square tests for independence should not be used because of the within-subject dependence among responses. An intuitively constructed version of the Pearson statistic is proposed to perform the test using bootstrap procedures to approximate its sampling distribution. First- and second-order adjustments to the proposed statistic are given in order to use a chi-square distribution approximation. A Bonferroni adjustment is proposed to perform the test when the joint set of responses for individual subjects is unavailable. Simulations show that the bootstrap procedures hold the correct size more consistently than the other procedures.

KEY WORDS: Bootstrap; Correlated binary data; Pearson statistic; Pick any/c; Simultaneous pairwise marginal independence.

1. Introduction

What types of cars do you own? What are your sources of veterinary information? For what criminal offenses have you been arrested? These are all example questions appearing on surveys where the respondent is prompted to pick any number of responses from a set of predefined items (Smith, Smith, and Noma, 1986; Umesh, 1995; Loughin and Scherer, 1998). Variables that summarize this type of "pick any" survey data have been called multiple-response (or pick any/c) categorical variables. Survey data arising from questions of this type present a unique challenge for analysis because of the dependence among responses provided by individual subjects.

Testing for independence between two categorical variables is often of interest. When at least one of the categorical variables can have multiple responses, traditional Pearson chi-square tests for independence should not be used because of the within-subject dependence among responses. Furthermore, a special kind of independence, called marginal independence, becomes of interest in the presence of multiple-response categorical variables. The purpose of this article is to develop new approaches to the testing of marginal independence between two multiple-response categorical variables. Agresti and Liu (1999) call this a test for simultaneous pairwise marginal independence (SPMI). The proposed tests are extensions to the traditional Pearson chi-square tests for independence testing between single-response categorical variables.

Developing methods to test for marginal independence is becoming increasingly important. Until recently, survey respondents answering questions about race have been forced to identify themselves according to a single race, in conflict with the increasing multicultural status of today's society. Permitting multiple race classifications, as the U.S. Census began doing in 2000, allows people to more accurately describe themselves. Tests comparing probabilities of another categorical variable across races must account for these multiple responses.

A second example comes from Loughin (1998) and Agresti and Liu (1999), who discuss a survey conducted by the Department of Animal Sciences at Kansas State University. Two questions in the survey asked Kansas farmers about their sources of veterinary information and their swine waste storage methods. For these questions, the farmers were permitted

Table 1

Marginal table for Kansas farmer data. There are a total of 279 farmers who participated in the survey. The percentage of farmers picking a source of veterinary information and waste storage method pair are given next to the counts. The shaded cell corresponds to the example given in Section 1.

	Sources of veterinary information									
	Professional consultant		Veterinarian		State or local extension service		Magazines		Feed companies & representatives	
Waste storage methods										
Lagoon	34	12.19%	54	19.35%	50	17.92%	63	22.58%	41	14.70%
Pit	17	6.09%	33	11.83%	34	12.19%	43	15.41%	37	13.26%
Natural drainage	6	2.15%	23	8.24%	30	10.75%	49	17.56%	34	12.19%
Holding tank	1	0.36%	4	1.43%	4	1.43%	6	2.15%	2	0.72%

to select as many responses as applied from a list of items. Table 1 summarizes the data in a 4 × 5 table. For example, 34 farmers picked professional consultant as a source of veterinary information and lagoon as a waste storage method. A researcher may be interested in determining whether sources of veterinary information are independent of waste storage methods in a similar manner as would be done in a traditional Pearson chi-square test applied to a contingency table with single-response categorical variables. The traditional Pearson chi-square test should not be used here because of the multiple responses. Instead, a test for SPMI can be performed to determine whether each source of veterinary information is simultaneously independent of each swine waste storage method. More specifically, 4 × 5 = 20 different 2 × 2 tables can be formed to marginally summarize all possible responses to item pairs. Table 2 shows the 2 × 2 table for professional consultant and lagoon. Independence is tested in each of the 20 2 × 2 tables simultaneously for a test of SPMI. The test is marginal because responses are summed over the other item choices for each of the multiple-response categorical variables. If SPMI is rejected, examination of the individual 2 × 2 tables can follow to determine why the rejection occurs. This is analogous to the F-protected t-test procedure that is often used in analysis of variance.

Tests for marginal independence have been only recently proposed in the presence of one multiple-response categorical variable. Bilder, Loughin, and Nettleton (2000) review the testing methods for a test of multiple marginal independence (MMI) between one single-response and one multiple-response categorical variable. They recommend bootstrapping a naive-sum (also called naive chi-squared) statistic proposed by Agresti and Liu (1999), performing bootstrap p-value combination methods, or using Bonferroni adjustments. Bilder and Loughin (2002) examine a test for conditional multiple marginal independence (CMMI), where the conditioning is with respect to a third single-response categorical variable. Similar conclusions are reached for the CMMI testing problem as for testing MMI.

Table 2

Professional consultant and lagoon 2 × 2 table. A "1" denotes a farmer picked that item and a "0" denotes the farmer did not pick that item. The shaded cell corresponds to the example given in Section 1.

		Professional consultant	
		1	0
Lagoon	1	34	109
	0	10	126

Little research has been done on testing for SPMI. Loughin (1998) suggests bootstrapping a Pearson statistic to perform the test. Thomas and Decady (2000) propose adjustments to this statistic in order to use a chi-square approximation to the sampling distribution. Both testing procedures have a defect that the proposed statistics are not invariant to the arbitrary coding of a 1 or 0 to denote whether or not a subject picks a particular item. This is discussed more in Section 3. Agresti and Liu (1999, 2001) suggest generalized log-linear models and multivariate binomial logit–normal models to test for SPMI. These and other model-based approaches are the subject of future research as discussed in Section 6.

The article is organized as follows. Section 2 presents the notation to be used and specifically defines the SPMI hypothesis. Section 3 proposes SPMI testing procedures that are extensions of a Pearson chi-square test statistic. Section 4 describes the application of the SPMI testing methods to the Kansas farmer data. Section 5 discusses a simulation study that examines size and power for the proposed tests. Section 6 gives concluding comments and recommendations.

2. Notation

Let W and Y denote the multiple-response categorical variables for an $r \times c$ table's row and column variables, respectively. Corresponding to the data in Table 1, sources of veterinary information are denoted by Y and waste storage methods are denoted by W. The categories for each multiple-response categorical variable are called items (Agresti and Liu, 1999). For example, lagoon is one of the items for waste storage method. Suppose W has r items and Y has c items. Also, suppose n subjects are sampled at random. Let $W_{si} = 1$ if a positive response is given for item i by subject s for $i = 1, \ldots, r$ and $s = 1, \ldots, n$; $W_{si} = 0$ for a negative response. Let Y_{sj} for $j = 1, \ldots, c$ and $s = 1, \ldots, n$ be similarly defined. The abbreviated notation, W_i and Y_j, refers generally to the binary response random variable for item i and j, respectively.

Table 3

Joint table for Kansas farmer data. The Y_j and W_i items correspond to the same ordering of the column and row items listed in Table 1. For example, Y_1 denotes professional consultant and W_1 denotes lagoon. The shaded region corresponds to the example given in Section 2.

			Y₁	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
			Y₂	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
			Y₃	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
			Y₄	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
			Y₅	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
0	**0**	**0**	**0**	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	**0**	**0**	**1**	0	2	3	0	2	0	1	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	**0**	**1**	**0**	2	12	19	7	12	0	1	0	4	3	0	0	0	0	3	5	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	**0**	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	**1**	**0**	**0**	4	7	7	2	4	0	6	2	4	2	1	1	0	1	0	3	3	0	0	0	0	0	0	0	0	0	0	0	0	0	1	3
0	**1**	**0**	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	**1**	**1**	**0**	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
0	**1**	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	**0**	**0**	**0**	11	9	22	0	10	3	2	2	13	1	1	3	2	0	3	3	15	0	0	0	1	0	0	0	1	0	0	0	1	0	2	3
1	**0**	**0**	**1**	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	**0**	**1**	**0**	0	0	2	0	0	0	0	0	1	0	0	0	0	0	1	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1
1	**0**	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	**1**	**0**	**0**	0	2	1	0	2	0	1	0	2	4	0	0	0	0	0	2	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	4
1	**1**	**0**	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
1	**1**	**1**	**0**	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
1	**1**	**1**	**1**	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
W₁	**W₂**	**W₃**	**W₄**																																

The set of correlated binary item responses for subject s are $\mathbf{Y}_s = (Y_{s1}, Y_{s2}, \ldots, Y_{sc})'$ and $\mathbf{W}_s = (W_{s1}, W_{s2}, \ldots, W_{sr})'$.

A "joint table" gives the cross-classification of responses to each possible set of item responses for W and Y. This is similar to the joint table described in Bilder et al. (2000) and the "expanded" and "complete" table described in Loughin and Scherer (1998) and Agresti and Liu (1999), respectively. Table 3 gives the joint table for the Kansas farmer data. For example, 15 farmers picked professional consultant as their only source of veterinary information and lagoon as their only waste storage method. Cell counts in the joint table are denoted by n_{gh} for the gth possible $(W_1, \ldots, W_r)$ and hth possible $(Y_1, \ldots, Y_c)$. The corresponding probability is denoted by τ_{gh}. Multinomial sampling is assumed to occur within the entire joint table; thus, $\sum_{g,h} \tau_{gh} = 1$.

Sparseness is usually the norm for joint tables. The number of cells in the joint table is 2^{r+c}, which can be quite large even for small values of r and c. For the Kansas farmer data example, there are $2^9 = 512$ cells with 434 of them zero. This table sparseness can have a detrimental effect on model-based testing approaches that need to estimate all τ_{gh} from the joint table. More details about sparseness are given in Section 6.

Let m_{ij} denote the number of observed positive responses to W_i and Y_j. A table summarizing these responses is called a marginal table because each m_{ij} is a sum of positive responses to items W_i and Y_j only, $m_{ij} = \sum_{\{g,h: W_i=1 \& Y_j=1\}} n_{gh}$. Table 1 is an example of a marginal table. The cells shaded in the body of Table 3 illustrate how joint table cell counts are summed to find the marginal count of 34 in Table 1. The marginal probability of a positive response to W_i and Y_j is denoted by π_{ij} and its maximum likelihood estimate (MLE) is $\hat{\pi}_{ij} = m_{ij}/n$.

The hypotheses for a test of SPMI are

$$\mathrm{H_o}\colon \pi_{ij} = \pi_{i\bullet}\pi_{\bullet j} \text{ for } i = 1, \ldots, r \text{ and } j = 1, \ldots, c,$$
$$\mathrm{H_a}\colon \text{At least one equality does not hold,}$$

where $\pi_{ij} = P(W_i = 1, Y_j = 1)$, $\pi_{i\bullet} = P(W_i = 1)$, and $\pi_{\bullet j} = P(Y_j = 1)$. This specifies marginal independence between each W_i and Y_j pair. The hypotheses can also be written another way. Consider the rc 2×2 pairwise item response tables formed for each W_i and Y_j pair (analogous to Table 2), and suppose the cells contain probabilities for each W_i and Y_j pair; i.e., $P(W_i = 1, Y_j = 1) = \pi_{ij}$, $P(W_i = 1, Y_j = 0) = \pi_{i\bullet} - \pi_{ij}$, $P(W_i = 0, Y_j = 1) = \pi_{\bullet j} - \pi_{ij}$, and $P(W_i = 0, Y_j = 0) = 1 - \pi_{i\bullet} - \pi_{\bullet j} + \pi_{ij}$. Provided none of these cells have 0 probability, SPMI can be written as $\mathrm{OR}_{WY,ij} = 1$ for $i = 1, \ldots, r$ and $j = 1, \ldots, c$ where OR is the abbreviation for odds ratio and $\mathrm{OR}_{WY,ij} = \pi_{ij}(1 - \pi_{i\bullet} - \pi_{\bullet j} + \pi_{ij})/[(\pi_{i\bullet} - \pi_{ij})(\pi_{\bullet j} - \pi_{ij})]$. Therefore, SPMI represents simultaneous independence in the rc 2×2 pairwise item response tables formed for each W_i and Y_j pair. The MLE for $\pi_{i\bullet}$ and $\pi_{\bullet j}$ are $\hat{\pi}_i = m_{i\bullet/n}$ and $\hat{\pi}_j = m_{\bullet j/n}$, respectively, where $m_{i\bullet} = \sum_{\{g: W_i=1\}} n_{gh}$ and $m_{\bullet j} = \sum_{\{h: Y_j=1\}} n_{gh}$.

Let $\tau_{g\bullet}$ be the probability of observing the gth possible $\mathbf{W}_s$ and $\tau_{\bullet h}$ denote the probability of observing the hth possible $\mathbf{Y}_s$. Joint independence is defined as $\tau_{g\bullet}\tau_{\bullet h} = \tau_{gh}$ for all g and h. This is a special case of SPMI; however, SPMI can exist without joint independence. With regard to the Kansas farmer data, joint independence indicates that each possible combination of the waste storage methods (a new 2^r-level single-response categorical variable) is independent of each possible combination of sources of veterinary information (a new 2^c-level single-response categorical variable). Often, however, the combinations of items are not of interest. Rather, the individual items are of interest. The less restrictive hypothesis of SPMI specifies that each waste storage method is pairwise independent of each source of veterinary information. This is a direct extension of the usual independence hypothesis in the single-response categorical variable case, which states that the probability for each pairwise response combination can be expressed as the product of the two marginal probabilities. Furthermore, SPMI is implied by joint independence but can exist without it. Thus, separate procedures are needed that can test this distinct hypothesis. For a further discussion on the differences between marginal and joint independence see Agresti and Liu (1999), Bilder et al. (2000), and Bilder and Loughin (2002).

3. Modified Pearson Based Statistics

Loughin (1998) suggests an intuitively chosen Pearson statistic,

$$X_M^2 = n\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(\hat{\pi}_{ij}-\hat{\pi}_{i\bullet}\hat{\pi}_{\bullet j})^2}{\hat{\pi}_{i\bullet}\hat{\pi}_{\bullet j}}$$

to test for SPMI. Unfortunately, this statistic is not invariant to the arbitrary designation of a 0 or 1 to a positive response. The statistic focuses only on the $W_i = 1$ and $Y_j = 1$ cells of the rc 2×2 pairwise item response tables. Thus, calculating X_M^2 using the $W_i = 1$ and $Y_j = 1$ counts (denote as $X_{M,1,1}^2$) may lead to a different observed value than, say, $X_{M,0,1}^2$ based on the $W_i = 0$ and $Y_j = 1$ counts. This means that one could reach different conclusions about SPMI depending on the coding of the data. For instance, with the Kansas farmer data, the four possible combinations of the W_i and Y_j $0-1$ responses result in values of $X_{M,1,1}^2 = 28.27$, $X_{M,1,0}^2 = 11.52$, $X_{M,0,1}^2 = 16.44$, and $X_{M,0,0}^2 = 6.08$. Similar to Agresti and Liu's (1999) proposal for the MMI problem, a solution is to calculate the statistic for all possible pairs of W_i and Y_j 0–1 responses. This produces four different statistics that can then be summed forming an invariant statistic,

$$X_S^2 = X_{M,1,1}^2 + X_{M,0,1}^2 + X_{M,1,0}^2 + X_{M,0,0}^2. \qquad (3.1)$$

This formulation of the X_S^2 statistic was first proposed by Bilder (2000).

An equivalent approach is to test for independence within each of the rc 2×2 item response tables using the Pearson chi-square test statistic. Denote the Pearson statistic for testing independence of W_i and Y_j as $X_{S,i,j}^2$. The resulting modified Pearson statistic for testing SPMI is

$$\begin{aligned}X_S^2 &= \sum_{i=1}^{r}\sum_{j=1}^{c}X_{S,i,j}^2\\ &= n\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(\hat{\pi}_{ij}-\hat{\pi}_{i\bullet}\hat{\pi}_{\bullet j})^2}{\hat{\pi}_{i\bullet}\hat{\pi}_{\bullet j}}\\ &\quad+\frac{[\hat{\pi}_{\bullet j}-\hat{\pi}_{ij}-\hat{\pi}_{\bullet j}(1-\hat{\pi}_{i\bullet})]^2}{\hat{\pi}_{\bullet j}(1-\hat{\pi}_{i\bullet})}\\ &\quad+\frac{[\hat{\pi}_{i\bullet}-\hat{\pi}_{ij}-\hat{\pi}_{i\bullet}(1-\hat{\pi}_{\bullet j})]^2}{\hat{\pi}_{i\bullet}(1-\hat{\pi}_{\bullet j})}\\ &\quad+\frac{[1-\hat{\pi}_{i\bullet}-\hat{\pi}_{\bullet j}+\hat{\pi}_{ij}-(1-\hat{\pi}_{i\bullet})(1-\hat{\pi}_{\bullet j})]^2}{(1-\hat{\pi}_{i\bullet})(1-\hat{\pi}_{\bullet j})}. \qquad (3.2)\end{aligned}$$

This formulation of the X_S^2 statistic was first proposed by Thomas and Decady (2000). Rearranging terms in (3.2) produces (3.1). Further, rearranging of terms in (3.2) produces the simplification,

$$X_S^2 = n\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(\hat{\pi}_{ij}-\hat{\pi}_{i\bullet}\hat{\pi}_{\bullet j})^2}{\hat{\pi}_{i\bullet}\hat{\pi}_{\bullet j}(1-\hat{\pi}_{i\bullet})(1-\hat{\pi}_{\bullet j})}.$$

If the rc Pearson statistics in X_S^2 are naively treated as independent, X_S^2 has an asymptotic χ_{rc}^2 distribution. SPMI is rejected if X_S^2 is greater than the $1-\alpha$ quantile of a χ_{rc}^2 distribution. In most cases, the rc Pearson statistics are not independent. Appendix A shows that $X_S^2 \xrightarrow{d} \sum_{p=1}^{rc}\lambda_p X_p^2$ under SPMI where the X_p^2's are independent χ_1^2 random variables and the λ_p are the eigenvalues of $\mathbf{D}^{-1}\mathbf{\Sigma}_0$ (defined in Appendix A). Bilder et al. (2000) and Bilder and Loughin (2002) discuss a variety of ways to approximate a similar asymptotic distribution for the MMI and CMMI tests. Extensions of these same approaches are outlined here.

3.1. *First- and Second-Order Adjustments*

Rao and Scott (1981) propose first-order adjustments to Pearson statistics and their sampling distributions in situations without simple random sampling. The first-order adjustment modifies the statistic to have the same asymptotic expectation as a χ^2 random variable. Bilder et al. (2000) use this adjustment to derive a test for MMI. Similar adjustments can also be made here to X_S^2 since its asymptotic distribution is a linear combination of independent χ_1^2 random variables. As mentioned in Thomas and Decady (2000) and shown here in Appendix B, the first-order adjustment for X_S^2 is $rc/\sum_{p=1}^{rc}\lambda_p = 1$. Thus, X_S^2 is "self-correcting." Thomas and Decady (2000) derive the first-order adjustment for $X_{M,1,1}^2$. Unfortunately, their adjusted statistic is not invariant to the arbitrary designation of a 0 or 1 to a positive response except in the extreme case of equality for the $\hat{\pi}_{i\bullet}$ for $i = 1, \ldots, r$ and equality for the $\hat{\pi}_{\bullet j}$ for $j = 1, \ldots, c$. Due to this lack of invariance, the first-order adjusted $X_{M,1,1}^2$ is not considered further in this article.

Bilder et al. (2000) and Bilder and Loughin (2001) show the first-order adjusted statistic often does not hold the correct size for problems involving only one multiple-response categorical variable. Because of these past problems, second-order adjustments to X_S^2 should be investigated to better approximate its sampling distribution. The second-order adjustment modifies the statistic to have the same asymptotic expectation and variance as a χ^2 random variable. The second-order adjusted statistic is $rcX_S^2/\sum_{p=1}^{rc}\lambda_p^2$, which can be approximated by a χ^2 random variable with $r^2c^2/\sum_{p=1}^{rc}\lambda_p^2$ degrees of freedom. Unfortunately, there is not a nice simplification for $\sum_{p=1}^{rc}\lambda_p^2$ as there is for $\sum_{p=1}^{rc}\lambda_p$ in the first-order adjustment.

Estimated eigenvalues are used to estimate the adjustment given above and these are based partly on the estimate of $\mathbf{\Sigma}_0$, the asymptotic covariance matrix of $(n)^{1/2}[\hat{\boldsymbol{\pi}} - \hat{\boldsymbol{\pi}}^R \otimes \hat{\boldsymbol{\pi}}^C]$ under SPMI (described in Appendix A). Note that $\mathbf{\Sigma}_0$ still depends on the $\{\tau_{gh}\}$. Finding estimates of these individual τ_{gh} under SPMI can be difficult because of the size and sparseness often encountered in the joint table. Specifically, Section 6 discusses difficulties with estimating the $\{\tau_{gh}\}$ for generalized loglinear models, and Bilder and Loughin (2002, p. 203) discuss problems with estimating the covariance matrix with relation to CMMI testing. Estimation of $\mathbf{\Sigma}$, the asymptotic covariance matrix without hypothesis restrictions, is somewhat less complicated and provides a consistent estimator of $\mathbf{\Sigma}_0$ under SPMI. Therefore, in the second-order adjusted X^2_S, the λ_p used are the eigenvalues for $\mathbf{D}^{-1}\mathbf{\Sigma}$ instead of $\mathbf{D}^{-1}\mathbf{\Sigma}_0$. The corresponding estimated matrices and eigenvalues can be found by substituting the estimators $\hat{\boldsymbol{\pi}}$ and $\hat{\boldsymbol{\tau}}$, as given in Appendix A, for $\boldsymbol{\pi}$ and $\boldsymbol{\tau}$, respectively.

3.2. *Nonparametric Bootstrap*

The sampling distribution of X^2_S can be approximated using a nonparametric bootstrap procedure. Although there are two multiple-response categorical variables here, the resampling is performed similarly to that used for testing MMI and CMMI (Bilder et al., 2000; Bilder and Loughin, 2002). To resample under independence of W and Y, $\mathbf{W}_s$ and $\mathbf{Y}_s$ are independently resampled with replacement from the data set. The test statistic calculated for the bth resample of size n is denoted by $X^{2*}_{S,b}$. The p-value is calculated as $B^{-1}\sum_b I(X^{2*}_{S,b} \geq X^2_S)$ where B is the number of resamples taken and $I()$ is the indicator function.

This procedure is actually resampling under joint independence, a special case of SPMI. When the similar form of resampling is done for the MMI and CMMI testing problems in Bilder et al. (2000, p. 1305) and Bilder and Loughin (2002, p. 205), the size of the test is not adversely affected. A simulation study discussed in Section 5.1 shows similar findings for the SPMI testing problem.

3.3. *Bootstrap p-Value Combination Methods*

Each $X^2_{S,i,j}$ gives a test for independence between each W_i and Y_j pair for $i = 1, \ldots, r$ and $j = 1, \ldots, c$. The p-values from each of these tests (using a χ^2_1 approximation) can be combined to form a new statistic, $\tilde{p}$. Combination methods previously used for the MMI and CMMI tests are the product of the p-values and the minimum of the p-values. Since the rc different tests are likely to be correlated, the usual p-value combination methods based on the independence of the p-values (see Hedges and Olkin, 1985) are not appropriate. The bootstrap can be used to approximate the sampling distribution of $\tilde{p}$ and a test can be developed. Resamples for the bootstrap procedure are taken the same way as described in Section 3.2. The p-value for the combined test is calculated as $B^{-1}\sum_b I(\tilde{p}^*_b \leq \tilde{p})$, where $\tilde{p}^*_b$ is the combined p-value calculated for the bth resample.

3.4. *Bonferroni Adjustment*

As an alternative to the bootstrap procedures, a Bonferroni adjustment can be applied to the rc $X^2_{S,i,j}$. SPMI is rejected if any $X^2_{S,i,j}$ is greater than the $1 - \alpha/(rc)$ quantile of a χ^2_1 distribution. A Bonferroni adjusted p-value can also be calculated by multiplying the minimum of the rc p-values by rc. The advantage of a Bonferroni adjustment approach is that it can be calculated without knowing the joint table of responses. The disadvantage of this approach is that for moderate to large r and c values, the Bonferroni adjustment to the critical value may be severe leading to a conservative test.

Table 4
SPMI testing method p-values for the Kansas farmer data. There are 10,000 *resamples used for the bootstrap methods.*

Method	Section	p-value
X^2_S with a χ^2_{rc} approximation	3 and 3.1	3.11×10^{-6}
Second-order adjusted X^2_S	3.1	3.07×10^{-5}
Bootstrap X^2_S	3.2	<0.0001
Bootstrap product of p-values	3.3	0.0001
Bootstrap minimum p-value	3.3	0.0034
Bonferroni adjustment	3.4	0.0037

3.5. *Follow-Up Analysis*

The previous subsections propose ways to test for SPMI. If SPMI is rejected, one would want to know why it is rejected. Since X^2_S is written in (3.2) as the sum of rc different Pearson chi-square test statistics, each $X^2_{S,i,j}$ can be used to measure why SPMI is rejected. The individual tests can be done using an asymptotic χ^2_1 approximation or the estimated sampling distribution of the individual statistics calculated in the proposed bootstrap procedures. A similar follow-up procedure is often used in analysis of variance. After an overall F-test for differences between treatment means is rejected, multiple comparison procedures are used to determine which mean pairs are different.

4. Application to the Kansas Farmer Data

The testing procedures of Section 3 are applied to the Kansas farmer data and the corresponding p-values are shown in Table 4. All methods indicate strong evidence against SPMI. Using the follow-up analysis approach outlined in Section 3.5, the $X^2_{S,i,j}$ and the corresponding p-values using χ^2_1 approximation are calculated. The significant pairwise combinations are (W_1, Y_1), (W_1, Y_2), (W_2, Y_2), (W_2, Y_5), (W_3, Y_1), and (W_3, Y_4) at the 0.05 significance level. If a Bonferroni adjusted significance level of $0.05/20 = 0.0025$ is used instead, only $(W_1, Y_1) =$ (Lagoon, Professional consultant) has a smaller p-value.

5. Simulation Study

A simulation study is performed to determine which testing procedures of Section 3 hold the correct size under a range of different situations and have power to detect various alternative hypotheses. The algorithm of Gange (1995) is used to simulate 500 data sets for each simulation setting investigated. For each simulated data set, the SPMI testing methods are applied (bootstrap methods use $B = 1000$), and for each method the proportion of data sets are recorded for which SPMI is rejected at the 0.05 nominal level. More specific details about how the simulation study is conducted and additional results are available from the first author.

5.1. *Type I Error*

To simulate data under SPMI, the $\mathrm{OR}_{WY,ij}$ are set to 1 for $i = 1, \ldots, r$ and $j = 1, \ldots, c$. Odds ratios between each W_i and $W_{i'}$ pair $(i < i')$ and each Y_j and $Y_{j'}$ pair $(j < j')$ are controlled as well, but not necessarily at a level of 1. These odds ratios are calculated as

$$\mathrm{OR}_{W,ii'} = \frac{P(W_i = 1 \text{ and } W_{i'} = 1)/P(W_i = 1 \text{ and } W_{i'} = 0)}{P(W_i = 0 \text{ and } W_{i'} = 1)/P(W_i = 0 \text{ and } W_{i'} = 0)},$$

for each W_i and $W_{i'}$ pair and in a similar manner for each Y_j and $Y_{j'}$ pair to form $\mathrm{OR}_{Y,jj'}$. The $\mathrm{OR}_{W,ii'}$ and $\mathrm{OR}_{Y,jj'}$ are set at values of 2 and 25 in the simulations to represent weak and strong pairwise dependence. Although observed strong pairwise dependence does not occur between items of the same multiple-response categorical variable for the Kansas farmer data, it can occur in practice. For example, the urinary tract infection data set described in Bilder and Loughin (2002) exhibits strong observed pairwise associations between the use of oral contraceptives and other contraceptive methods.

Table 5 shows the estimated type I error rates for 2 × 2 and 5 × 5 marginal table simulations. The 95% expected range of estimated type I error rates for testing methods holding the correct size is $0.05 \pm 2(0.05(1-0.05)/500)^{1/2} = (0.0305, 0.0695)$. All of the bootstrap methods are generally holding the correct size. The Bonferroni adjustment and second-order adjusted X_S^2 hold the correct size most of the time, but are too conservative sometimes. The X_S^2 statistic with a χ^2_{rc} approximation to its sampling distribution (first-order adjusted X_S^2) mostly holds the correct size for the $\mathrm{OR}_{W,ii'} = \mathrm{OR}_{Y,jj'} = 2$ simulations, but rejects too often when an odds ratio of 25 is present. This is because the procedure naively assumes the independence of the rc $X^2_{S,i,j}$. These results, regarding the χ^2_{rc} approximation to X_S^2, seem to contradict the simulation results in Thomas and Decady (2000). This is presumably because they generated their item responses using $\mathrm{OR}_{W,ii'}$ and $\mathrm{OR}_{Y,jj'}$ values all close to one (no measure of pairwise dependence is given in this article).

The results shown in Table 5 are based upon data generated under joint independence. Additional simulations (not shown) were performed under SPMI with joint independence violated. Despite resampling under joint independence, the test size was not adversely affected for the bootstrap methods. Details of the simulations are available from the first author.

5.2. *Power*

A limited simulation study was performed to examine the power of the SPMI testing methods. The X_S^2 statistic with a χ^2_{rc} approximation to its sampling distribution is excluded since it does not hold the correct size when $\mathrm{OR}_{W,ii'}$ and $\mathrm{OR}_{Y,jj'}$ are large. The study results (not shown) indicate there is not one particular best method with respect to power. Similar to the results in Bilder and Loughin (2002), the power of the bootstrap p-value combination methods is directly related to the type of alternative hypothesis. The minimum p-value method has larger power than the product of the p-values method when deviations from SPMI occur for only a few W_i and Y_j item pairs. The reverse is true when all or most W_i and Y_j item pairs deviate from SPMI. The bootstrap X_S^2 method tends to have similar power to the bootstrap product of the p-values method because of their statistics' similar construction. An analogous relationship holds for the Bonferroni adjustment and minimum p-value method. The power for the second-order adjusted X_S^2 tends to fall between the powers for

Table 5

Estimated type I error rates for the 2 × 2 and 5 × 5 marginal table simulations. The marginal probabilities are $\boldsymbol{\pi}^R = (0.4, 0.5)'$ and $\boldsymbol{\pi}^C = (0.2, 0.3)'$ ($\boldsymbol{\pi}^R$ and $\boldsymbol{\pi}^C$ are defined in Appendix A) for the 2 × 2 and $\boldsymbol{\pi}^R = \boldsymbol{\pi}^C = (0.1, 0.2, 0.3, 0.4, 0.5)'$ for the 5 × 5. Shaded cells correspond to estimated type I error rates outside of the 95% expected range.

Marginal table	$\mathrm{OR}_{W,ii'} = \mathrm{OR}_{Y,jj'}$	n	X_S^2 with a χ^2_{rc} approximation	Second-order adjusted X_S^2	Bootstrap X_S^2	Bootstrap product of p-values	Bootstrap minimum p-value	Bonferroni adjusted
2 × 2	2	12	0.054	0.052	0.038	0.038	0.050	0.020
		25	0.052	0.056	0.048	0.042	0.050	0.030
		50	0.066	0.070	0.064	0.064	0.044	0.040
		100	0.056	0.054	0.054	0.058	0.052	0.048
	25	12	0.084	0.040	0.036	0.032	0.054	0.020
		25	0.096	0.064	0.056	0.054	0.060	0.032
		50	0.084	0.056	0.054	0.052	0.060	0.046
		100	0.086	0.054	0.056	0.056	0.044	0.038
5 × 5	2	50	0.086	0.020	0.054	0.048	0.060	0.060
		100	0.062	0.028	0.040	0.038	0.052	0.054
		300	0.050	0.040	0.042	0.042	0.040	0.026
		500	0.072	0.050	0.070	0.056	0.058	0.060
	25	50	0.146	0.050	0.056	0.058	0.058	0.050
		100	0.156	0.046	0.050	0.052	0.048	0.034
		300	0.132	0.050	0.044	0.042	0.054	0.032
		500	0.136	0.058	0.052	0.054	0.042	0.028
	2 & 25 mix	50	0.110	0.026	0.038	0.042	0.034	0.034
		100	0.108	0.040	0.046	0.046	0.032	0.028
		300	0.128	0.056	0.068	0.068	0.056	0.046
		500	0.106	0.050	0.048	0.052	0.044	0.042

the other testing methods. The specific results of the study are available from the first author.

6. Discussion

The SPMI testing methods outlined here are counterparts to the independence testing methods already developed for single-response categorical variables. While the bootstrap methods may be the most computationally intensive of the testing methods, they most consistently hold the correct size and have power to detect various alternative hypotheses. The Bonferroni adjustment and second-order adjusted X_S^2 provide simpler methods to test for SPMI although they can be conservative at times.

The hypotheses of interest and tests proposed here have broader applications than just among problems involving pick any data. As suggested by a referee, the SPMI null hypothesis specifies that each binary random variable in a group is independent of each binary random variable in another group, regardless of the context of the measurements.

Model-based approaches to testing SPMI are currently being studied. Agresti and Liu (1999, Section 4) suggest using an adaptation of their generalized loglinear models used to test MMI. Bilder et al. (2000) show the maximum likelihood parameter estimation procedure can have difficulty with reaching convergence in the MMI testing situation. This is due to sparseness in the joint table. Since the joint table is typically larger ($2^r \times 2^c$) for the SPMI testing situation, sparseness and model convergence are even more of an issue. For example, convergence was not obtained after 150 iterations for the model fit to the Kansas farmer data set.

Other possible models include the multivariate binomial logit–normal model of Coull and Agresti (2000; Agresti and Liu, 2001), the alternating logistic regression model of Carey, Zeger, and Diggle (1993), and simple weighted least squares estimation of a generalized loglinear model (suggested by a referee). All of these procedures are currently being studied and will be the focus of a future manuscript.

Acknowledgements

This work was supported in part by National Science Foundation grants SES-0207212 and SES-0233321. The suggestions by an anonymous referee helped make this manuscript much more focused. Part of the first author's work was completed at Oklahoma State University.

Résumé

Dans les enquêtes, on permet souvent aux répondants de choisir autant de réponses qu'il faut dans une question à choix multiples. Les variables catégorielles qui récapitulent ce genre de données d'enquêtes sont appelées en anglais ≪pick any/c≫ ou variables à choix multiples. Il est souvent intéressant de tester l'indépendance entre deux variables catégorielles. Si toutes les variables catégorielles correspondent à des réponses multiples, le test traditionnel d'indépendance du Khi2 de Pearson ne devrait pas être utilisé à cause de la dépendance intra-sujet des réponses. Pour réaliser le test, on propose une version construite intuitivement de la statistique de Pearson, en utilisant des procédures de bootstrap pour approximer la distribution d'échantillonnage. Des ajustements au premier et second ordre sont donnés, afin d'utiliser une approximation de la distribution du Khi2. Un ajustement de type Bonferroni est proposé pour réaliser le test quand on ne dispose pas de l'ensemble des réponses pour un sujet. Des simulations montrent que les procédures de bootstrap fournissent la taille correcte du test plus efficacement que les autres procédures.

References

Agresti, A. and Liu, I.-M. (1999). Modeling a categorical variable allowing arbitrarily many category choices. *Biometrics* **55,** 936–943.

Agresti, A. and Liu, I.-M. (2001). Strategies for modeling a categorical variable allowing multiple category choices. *Sociological Methods & Research* **29,** 403–434.

Bilder, C. R. (2000). Testing for marginal independence with pick any/c variables. Ph.D. Dissertation, Department of Statistics, Kansas State University, Manhattan.

Bilder, C. R. and Loughin, T. M. (2001). On the first-order Rao–Scott correction of the Umesh–Loughin–Scherer statistic. *Biometrics* **57,** 1253–1255.

Bilder, C. R. and Loughin, T. M. (2002). Testing for conditional multiple marginal independence. *Biometrics* **58,** 200–208.

Bilder, C. R., Loughin, T. M., and Nettleton, D. (2000). Multiple marginal independence testing for pick any/c variables. *Communications in Statistics: Simulation and Computation* **29,** 1285–1316.

Carey, V., Zeger, S. L., and Diggle, P. (1993). Modelling multivariate binary data with alternating logistic regressions. *Biometrika* **80,** 517–526.

Coull, B. A. and Agresti, A. (2000). Random effects modeling of multiple binomial responses using the multivariate binomial logit–normal distribution. *Biometrics* **56,** 73–80.

Gange, S. J. (1995). Generating multivariate categorical variates using the iterative proportional fitting algorithm. *The American Statistician* **49,** 134–138.

Hedges, L. V. and Olkin, L. (1985). *Statistical Methods for Meta-Analysis.* San Diego: Academic Press.

Loughin, T. M. (1998). *Testing for independence in contingency tables with multiple row and column responses.* Technical Report, Kansas State University, Department of Statistics, Manhattan.

Loughin, T. M. and Scherer, P. N. (1998). Testing for association in contingency tables with multiple column responses. *Biometrics* **54,** 630–637.

Mathai, A. M. and Provost, S. B. (1992). *Quadratic Forms in Random Variables.* New York: Marcel Dekker.

Rao, J. N. K. and Scott, A. J. (1981). The analysis of categorical data from complex surveys: Chi-squared tests for goodness of fit and independence in two-way tables. *Journal of the American Statistical Association* **76,** 221–230.

Smith, W. R., Smith, D. R., and Noma, E. (1986). The multidimensionality of crime: A comparison of techniques for scaling delinquent careers. *Journal of Quantitative Criminology* **2,** 329–353.

Thomas, D. R. and Decady, Y. J. (2000). Analyzing categorical data with multiple responses per subject. *Statistics Society of Canada Proceedings of the Survey Methods Section,* 121–130.

Umesh, U. N. (1995). Predicting nominal variable relationships with multiple responses. *Journal of Forecasting* **14,** 585–596.

Received September 2002. *Revised June* 2003.
Accepted July 2003.

Appendix A

Let $\mathbf{m} = (m_{11}, m_{12}, \ldots, m_{rc})'$ and $\mathbf{n} = (n_{11}, n_{12}, \ldots, n_{2^r2^c})'$. Also, let $\mathbf{G}$ be a $r \times 2^r$ matrix with columns containing all possible values of $(W_1, \ldots, W_r)'$, and let $\mathbf{H}$ be a $c \times 2^c$ matrix with columns containing all possible values of $(Y_1, \ldots, Y_c)'$. For example, the column headers in Table 3 form $\mathbf{H}$ for the sources of veterinary information multiple-response categorical variable. Then $(\mathbf{G} \otimes \mathbf{H})\mathbf{n} = \mathbf{m}$ where $\otimes$ denotes the Kronecker product. This can be written equivalently as $(\mathbf{G} \otimes \mathbf{H})\hat{\boldsymbol{\tau}} = \hat{\boldsymbol{\pi}}$ where $\hat{\boldsymbol{\tau}} = \mathbf{n}/n$ and $\hat{\boldsymbol{\pi}} = \mathbf{m}/n$.

Define $\hat{\boldsymbol{\pi}}^R = (\hat{\pi}_{1\bullet}, \ldots, \hat{\pi}_{r\bullet})'$ and $\hat{\boldsymbol{\pi}}^C = (\hat{\pi}_{\bullet 1}, \ldots, \hat{\pi}_{\bullet c})'$. X_S^2 can be rewritten as $n(\hat{\boldsymbol{\pi}} - \hat{\boldsymbol{\pi}}^R \otimes \hat{\boldsymbol{\pi}}^C)'\hat{\mathbf{D}}^{-1}(\hat{\boldsymbol{\pi}} - \hat{\boldsymbol{\pi}}^R \otimes \hat{\boldsymbol{\pi}}^C)$ where $\hat{\mathbf{D}} = \mathrm{Diag}[\hat{\pi}_{i\bullet}\hat{\pi}_{\bullet j}(1 - \hat{\pi}_{i\bullet})(1 - \hat{\pi}_{\bullet j})]$. Using the joint asymptotic normality of $\hat{\boldsymbol{\tau}}$ and the delta method, it can be shown that $(n)^{1/2}[\hat{\boldsymbol{\pi}} - \hat{\boldsymbol{\pi}}^R \otimes \hat{\boldsymbol{\pi}}^C - (\boldsymbol{\pi} - \boldsymbol{\pi}^R \otimes \boldsymbol{\pi}^C)] \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$ where $E(\hat{\boldsymbol{\pi}}) = \boldsymbol{\pi}$, $E(\hat{\boldsymbol{\pi}}^R) = \boldsymbol{\pi}^R$, and $E(\hat{\boldsymbol{\pi}}^C) = \boldsymbol{\pi}^C$. Note that $\boldsymbol{\Sigma} = \mathbf{F}[\mathrm{Diag}(\boldsymbol{\tau}) - \boldsymbol{\tau}\boldsymbol{\tau}']\ \mathbf{F}'$ where $\boldsymbol{\tau} = (\tau_{11}, \tau_{12}, \ldots, \tau_{2^r2^c})'$, $\mathbf{F} = \mathbf{G} \otimes \mathbf{H} - \boldsymbol{\pi}^R \otimes [\mathbf{H}(\mathbf{j}_{2^r}' \otimes \mathbf{I}_{2^c})] - [\mathbf{G}(\mathbf{I}_{2^r} \otimes \mathbf{j}_{2^c}')] \otimes \boldsymbol{\pi}^C$, $\mathbf{I}_a$ is an $a \times a$ identity matrix, and $\mathbf{j}_a$ is an $a \times 1$ vector of 1's. Under SPMI, $(n)^{1/2}[\hat{\boldsymbol{\pi}} - \hat{\boldsymbol{\pi}}^R \otimes \hat{\boldsymbol{\pi}}^C] \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_0)$ where $\boldsymbol{\Sigma}_0$ is a matrix dependent on the τ_{gh}'s restricted by SPMI.

Note that $\hat{\mathbf{D}} \xrightarrow{p} \mathbf{D} = \mathrm{Diag}[\pi_{i\bullet}\pi_{\bullet j}(1 - \pi_{i\bullet})(1 - \pi_{\bullet j})]$. Let $\mathbf{Z} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_0)$. From Mathai and Provost (1992), $X_S^2 \xrightarrow{d} \mathbf{Z}'\mathbf{D}^{-1}\mathbf{Z} \sim \sum_{p=1}^{rc} \lambda_p X_p^2$ under SPMI where the λ_p are the eigenvalues of $\mathbf{D}^{-1}\boldsymbol{\Sigma}_0$ and the X_p^2 are independent χ_1^2 random variables, $p = 1, \ldots, rc$.

Appendix B

To obtain the first-order corrected statistic, find a δ, such that $E[\delta \sum_{p=1}^{rc} \lambda_p X_p^2] = rc$ where λ_p and X_p^2 are defined in Appendix A. This results in $\delta = rc/\sum_{p=1}^{rc} \lambda_p$. Because $\sum_{p=1}^{rc} \lambda_p = tr(\mathbf{D}^{-1}\boldsymbol{\Sigma}_0)$ and $\mathbf{D}^{-1}$ is a diagonal matrix, only the diagonal elements of $\boldsymbol{\Sigma}_0$ need to be found. The diagonal elements of $\boldsymbol{\Sigma}_0$ are the asymptotic variance of $(n)^{1/2}(\hat{\pi}_{ij} - \hat{\pi}_{i\bullet}\hat{\pi}_{\bullet j})$ under SPMI. Note that $\pi_{ij} - \pi_{i\bullet}\pi_{\bullet j}$ is a function of $\boldsymbol{\tau}$, i.e., $f(\boldsymbol{\tau}) = (\mathbf{g}_i' \otimes \mathbf{h}_j')\boldsymbol{\tau} - [\mathbf{g}_i'(\mathbf{I}_{2^r} \otimes \mathbf{j}_{2^c}')\boldsymbol{\tau}][\mathbf{h}_j'(\mathbf{j}_{2^r}' \otimes \mathbf{I}_{2^c})\boldsymbol{\tau}]$ where $\mathbf{g}_i'$ denotes the ith row of $\mathbf{G}$ and $\mathbf{h}_j'$ denotes the jth row of $\mathbf{H}$.

Using the delta-method, the asymptotic variance of $(n)^{1/2}(\hat{\pi}_{ij} - \hat{\pi}_{i\bullet}\hat{\pi}_{\bullet j})$ is $\dot{f}(\boldsymbol{\tau})[\mathrm{Diag}(\boldsymbol{\tau}) - \boldsymbol{\tau}\boldsymbol{\tau}']\dot{f}(\boldsymbol{\tau})'$, where $\dot{f}(\boldsymbol{\tau})$ is a $1 \times 2^{r+c}$ vector of partial derivatives with respect to $\boldsymbol{\tau}$ and $\mathrm{Diag}(\boldsymbol{\tau}) - \boldsymbol{\tau}\boldsymbol{\tau}'$ is the asymptotic covariance matrix for $(n)^{1/2}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau})$. The vector, $\dot{f}(\boldsymbol{\tau})$, is

$$\frac{d}{d\boldsymbol{\tau}}\left\{(\mathbf{g}_i' \otimes \mathbf{h}_j')\boldsymbol{\tau} - [\mathbf{g}_i'(\mathbf{I}_{2^r} \otimes \mathbf{j}_{2^c}')\boldsymbol{\tau}][\mathbf{h}_j'(\mathbf{j}_{2^r}' \otimes \mathbf{I}_{2^c})\boldsymbol{\tau}]\right\}$$
$$= \mathbf{g}_i' \otimes \mathbf{h}_j' - \pi_{i\bullet}[\mathbf{h}_j'(\mathbf{j}_{2^r}' \otimes \mathbf{I}_{2^c})] - \pi_{\bullet j}[\mathbf{g}_i'(\mathbf{I}_{2^r} \otimes \mathbf{j}_{2^c}')].$$

Then $\dot{f}(\boldsymbol{\tau})\mathbf{V}\dot{f}(\boldsymbol{\tau})'$ becomes

$$\begin{aligned}&\left\{\mathbf{g}_i' \otimes \mathbf{h}_j' - \pi_{i\bullet}[\mathbf{h}_j'(\mathbf{j}_{2^r}' \otimes \mathbf{I}_{2^c})] - \pi_{\bullet j}[\mathbf{g}_i'(\mathbf{I}_{2^r} \otimes \mathbf{j}_{2^c}')]\right\}\\ &\quad\times \left\{\mathrm{Diag}(\boldsymbol{\tau}) - \boldsymbol{\tau}\boldsymbol{\tau}'\right\}\\ &\quad\times \left\{\mathbf{g}_i \otimes \mathbf{h}_j - \pi_{i\bullet}[(\mathbf{j}_{2^r} \otimes \mathbf{I}_{2^c})\mathbf{h}_j] - \pi_{\bullet j}[(\mathbf{I}_{2^r} \otimes \mathbf{j}_{2^c})\mathbf{g}_i]\right\}. \quad \text{(B.1)}\end{aligned}$$

The above expression simplifies using the relationships: $(\mathbf{g}_i' \otimes \mathbf{h}_j')\boldsymbol{\tau} = \pi_{ij}$, $[\mathbf{g}_i'(\mathbf{I}_{2^r} \otimes \mathbf{j}_{2^c}')\tau] = \pi_{i\bullet}$, and $[\mathbf{h}_j'(\mathbf{j}_{2^r}' \otimes \mathbf{I}_{2^c})\tau] = \pi_{\bullet j}$. After further simplification of using $\pi_{ij} = \pi_{i\bullet}\pi_{\bullet j}$ under SPMI, equation (B.1) becomes $\pi_{i\bullet}\pi_{\bullet j}(1 - \pi_{i\bullet})(1 - \pi_{\bullet j})$. Then $tr(\mathbf{D}^{-1}\boldsymbol{\Sigma}_0) = \sum_{p=1}^{rc}[\pi_{i\bullet}\pi_{\bullet j}(1 - \pi_{i\bullet})(1 - \pi_{\bullet j})]^{-1}\pi_{i\bullet}\pi_{\bullet j} \times (1 - \pi_{i\bullet})(1 - \pi_{\bullet j}) = rc$. Thus, $\delta = rc/\sum_{p=1}^{rc} \lambda_p = 1$.