Statistica Sinica: Supplement

FULLY EFFICIENT JOINT FRACTIONAL IMPUTATION FOR INCOMPLETE BIVARIATE ORDINAL RESPONSES

Xichen She and Changbao Wu

University of Waterloo

Supplementary Material

This supplementary document contains elaborate discussions on modeling techniques for complete ordinal responses and existing methods for handling missing ordinal observations. Regularity conditions, proof of Theorem 3.2, justification of the bootstrap variance estimator and additional plots of the power functions of tests are also presented.

S1 Odds Ratios and Regression Models for Ordinal Responses

Our paper considers scenarios where one or both ordinal responses contain missing observations and explore strategies for valid and efficient inferences on the joint and marginal probabilities, association measures and regression analysis. A popular example on association measures is a set of different types of ordinal odds ratios, including the local (θ_{rj}^{L}), the cumulative (θ_{rj}^{C})

and the global (θ^G_{rj}) odds ratios, defined respectively as

$$\theta^L_{rj} = \frac{\pi_{rj}\pi_{r+1,j+1}}{\pi_{r,j+1}\pi_{r+1,j}}\,,$$

$$\theta^C_{rj} \;=\; \frac{\left(\sum_{b\le j}\pi_{rb}\right)\left(\sum_{b>j}\pi_{r+1,b}\right)}{\left(\sum_{b>j}\pi_{rb}\right)\left(\sum_{b\le j}\pi_{r+1,b}\right)}\,,$$

and

$$\theta^G_{rj} \;=\; \frac{\left(\sum_{a\le r}\sum_{b\le j}\pi_{ab}\right)\left(\sum_{a>r}\sum_{b>j}\pi_{ab}\right)}{\left(\sum_{a\le r}\sum_{b>j}\pi_{ab}\right)\left(\sum_{a>r}\sum_{b\le j}\pi_{ab}\right)}\,.$$

Note that both θ^C_{rj} and θ^G_{rj} have incorporated the ordinality of the responses in the definition.

In the absence of missing observations, Dale (1986) considered a bivariate case where each ordinal response was assumed to follow a cumulative link model. He modelled the global cross-ratios with a set of log-linear models and estimated parameters by maximizing the full-likelihood. Ekholm et al. (2013) chose to model the dependence ratios because of the computational advantage for full-likelihood inference. Other examples of analyzing ordinal responses with GEE include Lipsitz et al. (1994) and Heagerty and Zeger (1996). Agresti (2010) presented a comprehensive review on other related regression models for ordinal responses.

S2 Existing Methods for Handling Incomplete Ordinal Responses

The PSA re-weighting method usually requires stronger assumptions on the missing data mechanism such as the "covariate-dependent missingness" (CDM) (Little (1995)): $(\boldsymbol{\delta} \perp \boldsymbol{y}) \mid \boldsymbol{x}$, which is a stronger version of MAR. Under the CDM assumption, let $a_{ts}(\boldsymbol{x}) = Pr(\delta_1 = t, \delta_2 = s \mid \boldsymbol{x})$, $t, s = 0, 1$ be the "joint" propensity scores. This leads to the "marginal" propensity scores $P(\delta_1 = t \mid \boldsymbol{x}) = a_{t0} + a_{t1}$ for $t = 0, 1$ and $P(\delta_2 = s \mid \boldsymbol{x}) = a_{0s} + a_{1s}$ for $s = 0, 1$. Logistic regression models are commonly used for the "marginal" and "joint" propensity scores:

$$\begin{aligned}
&\operatorname{logit}\big[a_{10}(\boldsymbol{x}) + a_{11}(\boldsymbol{x})\big] = h_1(\boldsymbol{x}; \boldsymbol{\phi}_1), \quad \operatorname{logit}\big[a_{01}(\boldsymbol{x}) + a_{11}(\boldsymbol{x})\big] = h_2(\boldsymbol{x}; \boldsymbol{\phi}_2), \\
&\operatorname{logit}\big[a_{11}(\boldsymbol{x})\big] = h_3(\boldsymbol{x}; \boldsymbol{\phi}_3), \qquad \text{(S2.1)}
\end{aligned}$$

where h_1, h_2 and h_3 are known functions with unknown parameters $\boldsymbol{\phi}_1$, $\boldsymbol{\phi}_2$ and $\boldsymbol{\phi}_3$. Alternatively, we can consider $a_{ts}(\boldsymbol{x})$ as probability mass functions of a multinomial response with four categories and impose a multinomial GLM through the baseline-category logit model:

$$\begin{aligned}
&\log\big[a_{00}(\boldsymbol{x})/a_{11}(\boldsymbol{x})\big] = h_1(\boldsymbol{x}; \boldsymbol{\phi}_1), \quad \log\big[a_{01}(\boldsymbol{x})/a_{11}(\boldsymbol{x})\big] = h_2(\boldsymbol{x}; \boldsymbol{\phi}_2), \\
&\log\big[a_{10}(\boldsymbol{x})/a_{11}(\boldsymbol{x})\big] = h_3(\boldsymbol{x}; \boldsymbol{\phi}_3), \qquad \text{(S2.2)}
\end{aligned}$$

where we used the same notation h_1, h_2, h_3 and $\boldsymbol{\phi}_1$, $\boldsymbol{\phi}_2$, $\boldsymbol{\phi}_3$ as in (S2.1) but they may have different forms under model (S2.2). Let $\hat{a}_{i,ts}$ be $a_{ts}(\boldsymbol{x}_i)$ evaluated at the estimated parameters $\hat{\boldsymbol{\phi}}_k$, $k = 1, 2, 3$ from a chosen model. The PSA estimators for the cell probabilities π_{rj} are given by $\hat{\pi}^{psa}_{rj} = n^{-1}\sum_{i=1}^n \delta_{i1}\delta_{i2}\ \hat{a}^{-1}_{i,11}\boldsymbol{I}(y_{i1} = r, y_{i2} = j)$ and those for the marginal probabilities π_{r+} are given by $\hat{\pi}^{psa}_{r+} = n^{-1}\sum_{i=1}^n \delta_{i1}(\hat{a}_{i,10} + \hat{a}_{i,11})^{-1}\boldsymbol{I}(y_{i1} = r)$. The association parameter γ can be estimated by

$$\hat{\gamma}^{psa} = \big(C^{psa} - D^{psa}\big)/\big(C^{psa} + D^{psa}\big)\,, \tag{S2.3}$$

where $C^{psa} = 2\sum_{r<k}\sum_{j<l}\hat{\pi}^{psa}_{rj}\hat{\pi}^{psa}_{kl}$ and $D^{psa} = 2\sum_{r<k}\sum_{j>l}\hat{\pi}^{psa}_{rj}\hat{\pi}^{psa}_{kl}$.

The SRMI method can be carried out through the following five steps:

(1) Specify a regression model for the ordinal response variable with smaller rate of missing values (assuming it is y_1) involving y_1 and the fully observed covariates $\boldsymbol{x}$ and a prior distribution for regression coefficients. Draw imputed values of y_1 from the posterior predictive distribution for missing observations of y_1.

(2) Specify a regression model for y_2 with both $\boldsymbol{x}$ and y_1 as covariates along with a prior distribution for the regression coefficients. The imputed values of y_1 obtained from the previous step are treated as if they are observed, and we generate imputed values for the missing observations

of y_2 based on the posterior predictive distribution. Steps (1) and (2) generate an initial imputed data set.

(3) Specify a regression model for y_1 with both $\boldsymbol{x}$ and y_2 as covariates along with a prior distribution for the regression coefficients. Update the imputed values for y_1 with draws from the posterior predictive distribution based on the assumed model and the "complete" data set obtained in the previous step, treating the imputed values for y_2 as if they are observed.

(4) Repeat Steps (2) and (3) for a pre-specified number of times or until certain stability criterion is met to obtain the first imputed data set.

(5) Repeat Steps (1)-(4) to obtain multiple imputed data sets.

The SRMI method attempts to take into account of the correlation structure of the multivariate responses through a sequence of conditional models. Statistical analysis of the multiple imputed data sets follows from the general framework proposed by Rubin (1987), with standard methods applied to each of the imputed data sets separately and final results obtained through Rubin's combining rule. The general procedure resembles the Markov Chain Monte Carlo technique but the explicit relationship between the two and theoretical properties of the method have yet to be developed (Kenward and Carpenter (2007)).

When one of the ordinal responses y_1 and y_2 serves as a predictor in the regression model required for the SRMI method and also the proposed method, there are two possible approaches: the first is to ignore the ordinality and use dummy variables; the second is to assign proper scores to each level and treat it as a regular discrete numeric variable. For most applications, the dummy variable approach is preferable (Royston (2009)).

S3 Regularity Conditions and Proof of Theorem 3.2

We first present major results and regularity conditions on consistency and asymptotic normality of the m-estimators. Theorem S.1 follows directly from Theorem 2.1 and Lemma 2.4 in Newey and McFadden (1994). Theorem S.2 is adapted from Theorem 5.41 in van der Vaart (2000).

Theorem S.1. *Let $(X_1, ..., X_n)$ be an independent sample from some distribution P and $G(x;\boldsymbol{\eta})$ be a fixed vector-valued function with parameter $\boldsymbol{\eta}$ taking values in the parameter space Θ. Denote*

$$\Psi(\boldsymbol{\eta}) = E\big[G(X;\boldsymbol{\eta})\big] \quad \textit{and} \quad \Psi_n(\boldsymbol{\eta}) = \frac{1}{n}\sum_{i=1}^{n} G(X_i;\boldsymbol{\eta})\,.$$

Suppose that the following regularity conditions hold:

A1. $\Psi(\boldsymbol{\eta}) = 0$ *has a unique root at* $\boldsymbol{\eta}_0$*;*

A2. Θ *is compact;*

A3. $G(X;\boldsymbol{\eta})$ *is continuous at each* $\boldsymbol{\eta} \in \Theta$ *with probability one;*

A4. There exists $H(x)$ *such that* $|G(x;\boldsymbol{\eta})| \leq H(x)$ *for all* $\boldsymbol{\eta}$, *and* $E[H(X)] < \infty$.

Then the sequence of estimators $\hat{\boldsymbol{\eta}}_n$ *satisfying* $\Psi_n(\hat{\boldsymbol{\eta}}_n) = 0$ *converges to* $\boldsymbol{\eta}_0$ *in probability.*

Theorem S.2. *Suppose that, in addition to conditions A1-A4, the following conditions also hold:*

A5. $\boldsymbol{\eta}_0$ *is an interior point of* Θ*;*

A6. $G(x;\boldsymbol{\eta})$ *is twice continuously differentiable for every* x*;*

A7. The second-order partial derivatives of $G(x;\boldsymbol{\eta})$ *satisfy*

$$\left| \frac{\partial^2 G(x;\boldsymbol{\eta})}{\partial \eta_i \partial \eta_j} \right| \leq G_0(x)$$

for some integrable function $G_0(x)$ *for every* $\boldsymbol{\eta}$ *in a neighbourhood of* $\boldsymbol{\eta}_0$*;*

A8. $\dot{\Psi}(\boldsymbol{\eta}_0) = E\big[\partial G(X;\boldsymbol{\eta}_0)/\partial \boldsymbol{\eta}\big]$ *exists and is non-singular.*

Then

$$\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0 = -\Big[\dot{\Psi}(\boldsymbol{\eta}_0)\Big]^{-1} \frac{1}{n} \sum_{i=1}^{n} G(X_i;\boldsymbol{\eta}_0) + o_p\big(n^{-1/2}\big) .$$

We have shown that $\hat{\pi}_{rj}^{fi}$ is the solution to (3.16) and $(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)$ is the solution to (3.18). It follows that $(\hat{\pi}_{rj}^{fi}, \hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2)$ can be viewed as the solution to the joint estimating equations

$$\mathbf{0} = \sum_{i=1}^{n} U_{rj}(\boldsymbol{y}_i, \boldsymbol{\delta}_i, \boldsymbol{x}_i; \pi_{rj}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \quad \text{and} \quad \mathbf{0} = \sum_{i=1}^{n} \boldsymbol{S}_{obs}(\boldsymbol{y}_i, \boldsymbol{\delta}_i, \boldsymbol{x}_i; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2).$$

Here $\boldsymbol{S}_{obs}(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ is the score function derived from the observed likelihood and hence can be easily shown to be unbiased estimating functions. To show that U_{rj} is also unbiased, we evaluate the expectation of U_{rj} at the true parameter values $(\pi_{rj0}, \boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{20})$, which leads to

$$\begin{aligned}
& E\big[U_{rj}(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{x}; \pi_{rj0}, \boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{20})\big] \\
= \; & E\big\{E\big[U_{rj}(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{x}; \pi_{rj0}, \boldsymbol{\theta}_{10}, \boldsymbol{\theta}_{20}) \mid \boldsymbol{x}\big]\big\} \\
= \; & E\Big\{P(\delta_1 = 1, \delta_2 = 1 \mid y_1 = r, y_2 = j, \boldsymbol{x}) P(y_1 = r, y_2 = j \mid \boldsymbol{x}) \\
& + P(\delta_1 = 1, \delta_2 = 0 \mid y_1 = r, \boldsymbol{x}) P(y_2 = j \mid y_1 = r, \boldsymbol{x}) P(y_1 = r \mid \boldsymbol{x}) \\
& + P(\delta_1 = 0, \delta_2 = 1 \mid y_2 = j, \boldsymbol{x}) P(y_1 = r \mid y_2 = j, \boldsymbol{x}) P(y_2 = j \mid \boldsymbol{x}) \\
& + P(\delta_1 = 0, \delta_2 = 0 \mid \boldsymbol{x}) P(y_1 = r, y_2 = j \mid \boldsymbol{x})\Big\} - \pi_{rj0} \\
= \; & E\Big\{P(y_1 = r, y_2 = j \mid \boldsymbol{x})\Big\} - \pi_{rj0} \\
= \; & 0\,.
\end{aligned}$$

Let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$, $\boldsymbol{\eta} = (\pi_{rj}, \boldsymbol{\theta}')'$ and $G(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{x}; \boldsymbol{\eta}) = (U_{rj}(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{x}; \pi_{rj}, \boldsymbol{\theta}), \boldsymbol{S}_{obs}'(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{x}; \boldsymbol{\theta}))'$. By Theorem S.1, the estimator $\hat{\pi}_{rj}^{fi}$ is consistent under regularity conditions $A1$-$A4$. If conditions $A5$-$A8$ also hold, by Theorem S.2, we further conclude that

$$\begin{pmatrix} \hat{\pi}_{rj}^{fi} - \pi_{rj0} \\ \hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \end{pmatrix} = -\left[\dot{\Psi}(\pi_{rj0}, \boldsymbol{\theta}_0)\right]^{-1} \begin{pmatrix} n^{-1}\sum_{i=1}^n U_{i,rj}(\pi_{rj0}, \boldsymbol{\theta}_0) \\ n^{-1}\sum_{i=1}^n \boldsymbol{S}_{i,obs}(\boldsymbol{\theta}_0) \end{pmatrix} + o_p\left(n^{-1/2}\right), \tag{S3.1}$$

where $\boldsymbol{\theta}_0 = (\boldsymbol{\theta}_{10}', \boldsymbol{\theta}_{20}')'$, $U_{i,rj}$, $\boldsymbol{S}_{i,obs}$ are functions U_{rj}, $\boldsymbol{S}_{obs}$ evaluated at the ith observation and

$$\dot{\Psi}(\pi_{rj0}, \boldsymbol{\theta}_0) = \begin{pmatrix} E[\partial U_{rj}/\partial \pi_{rj}] & E[\partial U_{rj}/\partial \boldsymbol{\theta}] \\ \mathbf{0} & E[\partial \boldsymbol{S}_{obs}/\partial \boldsymbol{\theta}] \end{pmatrix}_{\pi_{rj0}, \boldsymbol{\theta}_0}.$$

It is easy to see that $E\left[\partial U_{rj}/\partial \pi_{rj}\right] = -1$ and $E[\partial \boldsymbol{S}_{obs}/\partial \boldsymbol{\theta}]_{\boldsymbol{\theta}_0} = -\boldsymbol{I}_{obs}$. To find the expression for $E[\partial U_{rj}/\partial \boldsymbol{\theta}]$, we note that U_{rj} given in (3.15) depends on $\boldsymbol{\theta}$ through W defined in (3.5). For the second term of U_{rj}, we have

$$\begin{aligned} \partial W((y_1, j), (1, 0), \boldsymbol{x}; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)/\partial \boldsymbol{\theta} &= (\mathbf{0}, \partial P(y_2 = j \mid y_1, \boldsymbol{x}; \boldsymbol{\theta}_2)/\partial \boldsymbol{\theta}_2) \\ &= (\mathbf{0}, \boldsymbol{S}_2(j, y_1, \boldsymbol{x}; \boldsymbol{\theta}_2) P(y_2 = j \mid y_1, \boldsymbol{x}; \boldsymbol{\theta}_2)). \end{aligned}$$

For the third term of U_{rj}, we have

$$\begin{aligned}
& \partial W((r,y_2),(0,1),\boldsymbol{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)/\partial\boldsymbol{\theta} \\
= & \ \partial P(y_1=r\mid y_2,\boldsymbol{x};\boldsymbol{\theta})/\partial\boldsymbol{\theta} \\
= & \ P(y_1=r\mid y_2,\boldsymbol{x};\boldsymbol{\theta})\Big\{\partial\log\big[P(y_1=r\mid y_2,\boldsymbol{x};\boldsymbol{\theta})\big]/\partial\boldsymbol{\theta}\Big\} \\
= & \ P(y_1=r\mid y_2,\boldsymbol{x};\boldsymbol{\theta})\begin{pmatrix}\boldsymbol{S}_1(r,\boldsymbol{x};\boldsymbol{\theta}_1)-E[\boldsymbol{S}_1(y_1,\boldsymbol{x};\boldsymbol{\theta}_1)\mid y_2,\boldsymbol{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2] \\ \boldsymbol{S}_2(y_2,r,\boldsymbol{x};\boldsymbol{\theta}_2)-E[\boldsymbol{S}_2(y_2,y_1,\boldsymbol{x};\boldsymbol{\theta}_2)\mid y_2,\boldsymbol{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2]\end{pmatrix}',
\end{aligned}$$

where the last equation uses the conditional probability mass function of y_1 given y_2 in (3.12). Similarly, for the fourth term of U_{rj}, we have

$$\begin{aligned}
\partial W((r,j),(0,0),\boldsymbol{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)/\partial\boldsymbol{\theta} &= \partial P(y_1=r,y_2=j\mid\boldsymbol{x};\boldsymbol{\theta})/\partial\boldsymbol{\theta} \\
&= P(y_1=r,y_2=j\mid\boldsymbol{x};\boldsymbol{\theta})(\boldsymbol{S}_1'(r,\boldsymbol{x};\boldsymbol{\theta}_1),\boldsymbol{S}_2'(j,r,\boldsymbol{x};\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)).
\end{aligned}$$

After some tedious but straightforward algebra, it can be shown that

$$\begin{aligned}
E\big[\partial U_{rj}/\partial\boldsymbol{\theta}\big]_{\boldsymbol{\theta}_0} &= E\Big\{P(y_1=r,y_2=j\mid\boldsymbol{x};\boldsymbol{\theta}_0)\big[\boldsymbol{S}((r,j),\boldsymbol{x};\boldsymbol{\theta}_0)-\boldsymbol{S}_{obs}((r,j),\boldsymbol{\delta},\boldsymbol{x};\boldsymbol{\theta}_0)\big]'\Big\} \\
&= E\Big\{\boldsymbol{I}(y_1=r,y_2=j)\big[\boldsymbol{S}((r,j),\boldsymbol{x};\boldsymbol{\theta}_0)-\boldsymbol{S}_{obs}((r,j),\boldsymbol{\delta},\boldsymbol{x};\boldsymbol{\theta}_0)\big]'\Big\} \\
&= \kappa_{rj}\,,
\end{aligned}$$

where $\boldsymbol{S}$ and $\boldsymbol{S}_{obs}$ are defined in Theorem 3.2. By the inverse formula for block matrix, we have

$$-\Big[\dot{\Psi}(\pi_{rj0},\boldsymbol{\theta}_0)\Big]^{-1}=\begin{pmatrix}1 & \kappa_{rj}\boldsymbol{I}_{obs}^{-1} \\ \boldsymbol{0} & \boldsymbol{I}_{obs}^{-1}\end{pmatrix}.$$

The asymptotic variance of $\hat{\pi}_{rj}^{fi}$ follows from (S3.1) and the central limit theorem. Extending the results to the vector parameters $\hat{\boldsymbol{\pi}}^{fi}$ is straightforward.

S4 Justification of the Bootstrap Variance Estimator

It suffices to show that the resulting estimator $\hat{\boldsymbol{\pi}}^{(k)}$ based on the kth bootstrap sample is asymptotically equivalent to $\hat{\boldsymbol{\pi}}^{fi}$. To see this, consider an infinite path of the response variables $\mathcal{P} = \big\{(\boldsymbol{y}_1, \boldsymbol{\delta}_1, \boldsymbol{x}_1), \ldots, (\boldsymbol{y}_n, \boldsymbol{\delta}_n, \boldsymbol{x}_n), \ldots\big\}$. By the Strong Law of Large Numbers, the following two conditions hold for almost all paths.

$$
\begin{aligned}
&\text{(i)}.\ n^{-1}\sum_{i=1}^{n} \mathring{G}(\boldsymbol{y}_i, \boldsymbol{\delta}_i, \boldsymbol{x}_i; \hat{\boldsymbol{\eta}}) \longrightarrow E\Big[\mathring{G}(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{x}; \boldsymbol{\eta}_0)\Big]\,, \\
&\text{(ii)}.\ n^{-1}\sum_{i=1}^{n} G^{\otimes 2}(\boldsymbol{y}_i, \boldsymbol{\delta}_i, \boldsymbol{x}_i; \hat{\boldsymbol{\eta}}) \longrightarrow E\Big[G^{\otimes 2}(\boldsymbol{y}, \boldsymbol{\delta}, \boldsymbol{x}; \boldsymbol{\eta}_0)\Big]\,,
\end{aligned}
$$

where $\mathring{G} = \partial G/\partial \boldsymbol{\eta}$, $\hat{\boldsymbol{\eta}} = (\hat{\pi}_{rj}, \hat{\boldsymbol{\theta}}_1', \hat{\boldsymbol{\theta}}_2')$ and $\boldsymbol{\eta}_0 = (\pi_{rj0}, \boldsymbol{\theta}_{10}', \boldsymbol{\theta}_{20}')$. Conditional on one such path, the bootstrap samples form a triangular array as shown in Table 1, where the nth row consists of n *i.i.d.* samples from the empirical distribution of the first n points in the path, i.e., $\big\{(\boldsymbol{y}_1, \boldsymbol{\delta}_1, \boldsymbol{x}_1), \ldots, (\boldsymbol{y}_n, \boldsymbol{\delta}_n, \boldsymbol{x}_n)\big\}$. It then follows that

Table 1: The Triangular Array Formed by Bootstrap Samples

$$
\begin{array}{llll}
(\tilde{\boldsymbol{y}}_{11}^{(k)}, \tilde{\boldsymbol{\delta}}_{11}^{(k)}, \tilde{\boldsymbol{x}}_{11}^{(k)}) & & & \\
(\tilde{\boldsymbol{y}}_{21}^{(k)}, \tilde{\boldsymbol{\delta}}_{21}^{(k)}, \tilde{\boldsymbol{x}}_{21}^{(k)}) & (\tilde{\boldsymbol{y}}_{22}^{(k)}, \tilde{\boldsymbol{\delta}}_{22}^{(k)}, \tilde{\boldsymbol{x}}_{22}^{(k)}) & & \\
\vdots & \vdots & & \\
(\tilde{\boldsymbol{y}}_{n1}^{(k)}, \tilde{\boldsymbol{\delta}}_{n1}^{(k)}, \tilde{\boldsymbol{x}}_{n1}^{(k)}) & (\tilde{\boldsymbol{y}}_{n2}^{(k)}, \tilde{\boldsymbol{\delta}}_{n2}^{(k)}, \tilde{\boldsymbol{x}}_{n2}^{(k)}) & \cdots & (\tilde{\boldsymbol{y}}_{nn}^{(k)}, \tilde{\boldsymbol{\delta}}_{nn}^{(k)}, \tilde{\boldsymbol{x}}_{nn}^{(k)}) \\
\vdots & \vdots & & \vdots
\end{array}
$$

$$
\begin{aligned}
E\Big\{G(\tilde{\boldsymbol{y}}_{ni}^{(k)}, \tilde{\boldsymbol{\delta}}_{ni}^{(k)}, \tilde{\boldsymbol{x}}_{ni}^{(k)}; \hat{\boldsymbol{\eta}}) \mid \mathcal{P}\Big\} &= n^{-1} \sum_{i=1}^{n} G(\boldsymbol{y}_i, \boldsymbol{\delta}_i, \boldsymbol{x}_i; \hat{\boldsymbol{\eta}}) = \mathbf{0}\,, \\
Var\Big\{G(\tilde{\boldsymbol{y}}_{ni}^{(k)}, \tilde{\boldsymbol{\delta}}_{ni}^{(k)}, \tilde{\boldsymbol{x}}_{ni}^{(k)}; \hat{\boldsymbol{\eta}}) \mid \mathcal{P}\Big\} &= n^{-1} \sum_{i=1}^{n} G^{\otimes 2}(\boldsymbol{y}_i, \boldsymbol{\delta}_i, \boldsymbol{x}_i; \hat{\boldsymbol{\eta}})\,.
\end{aligned}
$$

The bootstrap estimator $\hat{\boldsymbol{\eta}}^{(k)} = (\hat{\pi}_{rj}^{(k)}, \hat{\boldsymbol{\theta}}_1^{(k)'}, \hat{\boldsymbol{\theta}}_2^{(k)'})$ solves the following estimating equations

$$
\mathbf{0} = n^{-1} \sum_{i=1}^{n} G(\tilde{\boldsymbol{y}}_{ni}^{(k)}, \tilde{\boldsymbol{\delta}}_{ni}^{(k)}, \tilde{\boldsymbol{x}}_{ni}^{(k)}; \boldsymbol{\eta})\,.
$$

If we expand $n^{-1} \sum_{i=1}^{n} G(\tilde{\boldsymbol{y}}_{ni}^{(k)}, \tilde{\boldsymbol{\delta}}_{ni}^{(k)}, \tilde{\boldsymbol{x}}_{ni}^{(k)}; \hat{\boldsymbol{\eta}}^{(k)})$ around $\hat{\boldsymbol{\eta}}$, it is not difficult to see that

$$
\begin{aligned}
\hat{\boldsymbol{\eta}}^{(k)} - \hat{\boldsymbol{\eta}} = &\left\{-n^{-1} \sum_{i=1}^{n} \mathring{G}(\tilde{\boldsymbol{y}}_{ni}^{(k)}, \tilde{\boldsymbol{\delta}}_{ni}^{(k)}, \tilde{\boldsymbol{x}}_{ni}^{(k)}; \hat{\boldsymbol{\eta}})\right\}^{-1} \left\{n^{-1} \sum_{i=1}^{n} G(\tilde{\boldsymbol{y}}_{ni}^{(k)}, \tilde{\boldsymbol{\delta}}_{ni}^{(k)}, \tilde{\boldsymbol{x}}_{ni}^{(k)}; \hat{\boldsymbol{\eta}})\right\} \\
&+ o_p(n^{-1/2})\,.
\end{aligned}
$$

By the Weak Law of Large Numbers and Central Limit Theorem for triangular arrays and also condition (i) and (ii), following the same partitioning arguments as in Section S3, we show that

$$n^{1/2}(\hat{\boldsymbol{\pi}}^{(k)} - \hat{\boldsymbol{\pi}}^{fi}) \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{\Sigma}^{fi})\,.$$

S5 Additional Plots of Power Functions of Tests

The power of a test is computed as the simulated rejection probability under the given scenario. Plots of the power function for missing pattern 5221 are shown in Figures 1 and 2, corresponding to sample sizes at $n = 200$ and 500. Each plot shows the power functions of three different tests: `JFI_non`, `JFI_nul` and SRMI. The first test uses the regular linearization variance estimator without considering the null hypothesis; the second test uses the linearization variance estimator under the null hypothesis (i.e., $\pi_{rj} = \pi_{r+}\pi_{+j}$); the third test uses the regular point and variance estimators for the SRMI method.

References

Agresti, A. (2010). *Analysis of Ordinal Categorical Data (2nd edition)*, Wiley, New York.

Dale, J. R. (1986). Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics*, **42**, 909–917.

Ekholm, A., Jokinen, J., McDonald, J. W. and Smith, P. W. (2013). Joint regression and association modeling of longitudinal ordinal data. *Biometircs*, **59**, 795–803.

Heagerty, P. J. and Zeger, S. L. (1996). Marginal regression models for clustered ordinal measurements. *Journal of the American Statistical Association*, **91**, 1024–1036.

Kenward, M. G. and Carpenter, J. (2007). Multiple imputation: Current perspectives. *Statistical Methods in Medical Research*, **16**, 199–218.

Lipsitz, S. R., Kim, K. and Zhao, L. (1994). Analysis of repeated categorical data using generalized estimating equations. *Statistics in Medicine*, **13**, 1149–1163.

Little, R. J. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, **90**, 1112–1121.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics*, **4** 2111–2245.

Royston, P. (2009). Multiple imputation of missing values: Further update of ice, with an emphasis on categorical variables. *Stata Journal*, **9**, 466–477.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*, Wiley Series in Probability and Statistics, Wiley.

van der Vaart, A. W. (2000). *Asymptotic Statistics*, Volume 3, Cambridge University Press, Cambridge.

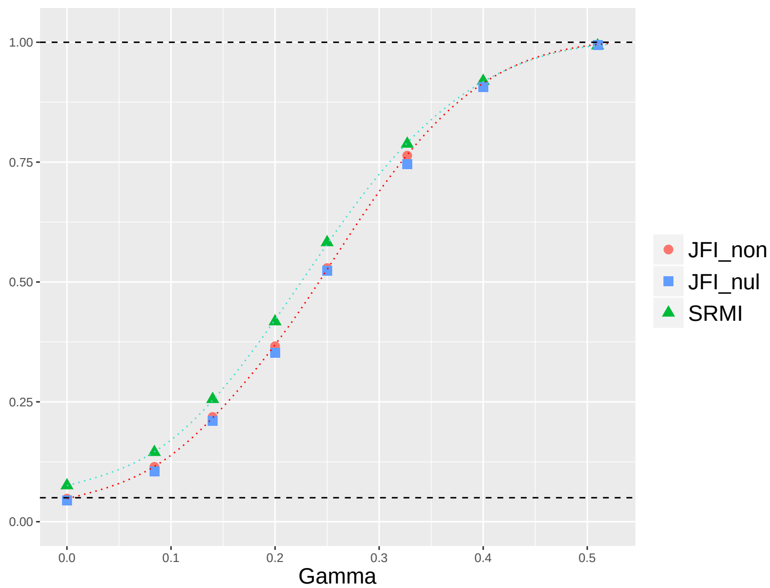

Figure 1: Power Function with $n = 200$ and Pattern 5221

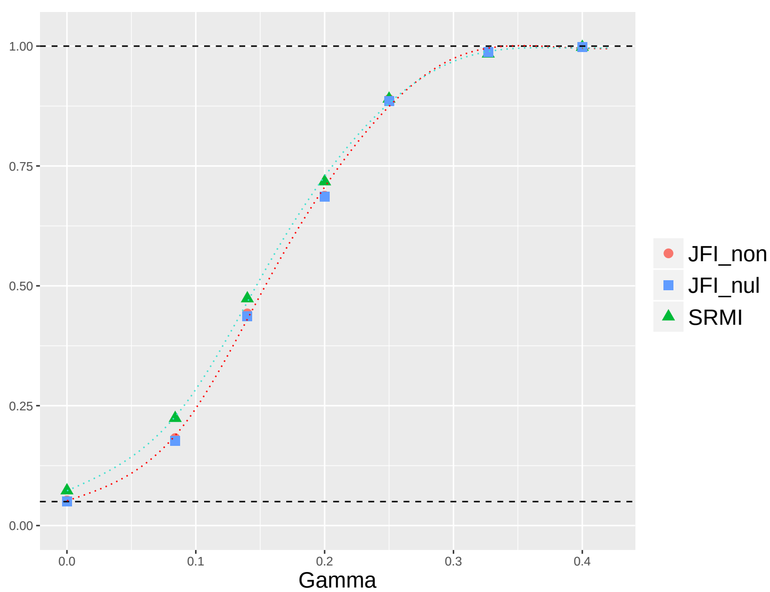

Figure 2: Power Function with $n = 500$ and Pattern 5221