A Note on Statistical Inference for Noisy Incomplete 1-Bit Matrix*

Yunxiao Chen¹, Chengcheng Li², and Gongjun Xu²

¹Department of Statistics, London School of Economics and Political Science
²Department of Statistics, University of Michigan

Abstract

We consider the statistical inference for noisy incomplete 1-bit matrix. Instead of observing a subset of real-valued entries of a matrix $M$, we only have one binary (1-bit) measurement for each entry in this subset, where the binary measurement follows a Bernoulli distribution whose success probability is determined by the value of the entry. Despite the importance of uncertainty quantification to matrix completion, most of the categorical matrix completion literature focus on point estimation and prediction. This paper moves one step further towards the statistical inference for 1-bit matrix completion. Under a popular nonlinear factor analysis model, we obtain a point estimator and derive its asymptotic distribution for any linear form of $M$ and latent factor scores. Moreover, our analysis adopts a flexible missing-entry design that does not require a random sampling scheme as required by most of the existing asymptotic results for matrix completion. The proposed estimator is statistically efficient and optimal, in the sense that the Cramer-Rao lower bound is achieved asymptotically for the model parameters. Two applications are considered, including (1) linking two forms of an educational test and (2) linking the roll call voting records from multiple years in the United States senate. The first application enables the comparison between examinees who took different test forms, and the second application allows us to compare the liberal-conservativeness of senators who did not serve in the senate at the same time.

Keywords: 1-bit matrix; Matrix completion; Binary data; Asymptotic normality; Nonlinear latent variable model.

*This research is partially supported by National Science Foundation CAREER SES-1846747 and Institute of Education Sciences R305D200015.
1 Introduction

Noisy low-rank matrix completion is concerned with the recovery of a low-rank matrix when only a fraction of noisy entries are observed. This topic has received much attention in the past decade, as a result of its vast applications in practical contexts such as collaborative filtering (Goldberg et al., 1992), system identification (Liu and Vandenberghe, 2010) and sensor localisation (Biswas et al., 2006). While the majority of the literature considers the completion of real-valued observations (Candès and Recht, 2009; Candès and Tao, 2010; Keshavan et al., 2010; Koltchinskii et al., 2011; Negahban and Wainwright, 2012; Chen et al., 2020a), many practical problems involve categorical-valued matrices, such as the famous Netflix challenge. Several works have been done on the completion of categorical matrix, including Davenport et al. (2014) and Bhaskar and Javanmard (2015) for 1-bit matrix, and Klopp et al. (2015) and Bhaskar (2016) for categorical matrix, whose entries can take multiple discrete values. In these works, low-dimensional nonlinear probabilistic models are assumed to handle the categorical data.

Despite the importance of uncertainty quantification to matrix completion, most of the matrix completion literature focus on point estimation and prediction, while statistical inference has received attention only recently. Specifically, Chen et al. (2019a) and Xia and Yuan (2021) considered statistical inference under the linear models and derived asymptotic normality results. The statistical inference for categorical matrices is more challenging due to the involvement of nonlinear models. To our best knowledge, no work has been done to provide statistical inference for the completion of categorical matrices. In addition to nonlinearity, another challenge in modern theoretical analysis of matrix completion concerns the double asymptotic regime where both the numbers of rows and columns are allowed to grow to infinity. Under this asymptotic regime, both the dimension of the parameters and the number of observable entries grow with the numbers of rows and columns. However, existing theory on the statistical inference for diverging number of parameters (Portnoy, 1988; He and Shao, 2000; Wang, 2011) is not directly applicable, as the dimension of the parameter space in the current problem grows faster than that is typically needed for asymptotic normality; see Section 3 for further discussions.

In this paper, we move one step further towards the statistical inference for the completion
of categorical matrix. Specifically, we consider the inference for 1-bit matrix completion under a unidimensional nonlinear factor analysis model with the logit link. Such a nonlinear factor model is one of the most popular models for multivariate binary data, having received much attention from the theoretical perspective (Andersen, 1970; Haberman, 1977; Lindsay et al., 1991; Rice, 2004), as well as wide applications in various areas, including educational testing (van der Linden and Hambleton, 2013), word acquisition analysis (Kidwell et al., 2011), syntactic comprehension (Gutman et al., 2011), and analysis of health outcomes (Hagquist and Andrich, 2017). It is also referred to as the Rasch model (Rasch, 1960) in psychometrics literature. Despite the popularity and extensive research of the model, its use to 1-bit matrix completion and related statistical inferences for the latent factors and model parameters have not been explored. The considered nonlinear factor model is also closely related to the Bradley-Terry model (Bradley and Terry, 1952; Simons and Yao, 1999; Han et al., 2020) for directed random graphs and the β-model (Chatterjee et al., 2011; Rinaldo et al., 2013) for undirected random graphs. In fact, the considered model can be viewed as a Bradley-Terry model or β-model for bipartite graphs (Rinaldo et al., 2013). However, the analysis of bipartite graphs is more involved, for which the results and proof strategies in the existing works no longer apply and new technical tools are needed.

Specifically, we introduce a likelihood-based estimator under the nonlinear factor analysis model for 1-bit matrix completion. Under a very flexible missing-entry setting that does not require a random sampling scheme, asymptotic normality results are established that allow us to draw statistical inference. These results suggest that our estimator is asymptotically efficient and optimal, in the sense that the Cramer-Rao lower bound is achieved for model parameters. The proposed method and theory are applied to two real-world problems including (1) linking two forms of a college admission test that have common items and (2) linking the voting records from multiple years in the United States senate. In the first application, the proposed method allows us to answer the question “for examinees A and B who took different test forms, would examinee A perform significantly better than examinee B, if they had taken the same test form?” In the second application, it can answer the questions such as “Is Republican senator Marco Rubio significantly more conservative than Republican senator Judd Gregg?” Note that Marco Rubio and Judd Gregg had not served in the United States senate at the same time.
missingness in these applications does not satisfy the commonly assumed random sampling schemes for matrix completion.

The rest of the paper is organized as follows. In Section 2, we introduce the considered factor model and discuss its application to 1-bit matrix completion. In Section 3, we establish the asymptotic normality for the maximum likelihood estimator. A simulation study is given in Section 4 and two real-data applications are presented in Section 5. We conclude with discussions on the limitations of the current work and future directions in Section 6.

2 Model and Estimation

Let $Y$ be a 1-bit matrix with $N$ rows and $J$ columns and $Y_{ij} \in \{0, 1\}$ be the entries of $Y$, $i = 1, ..., N$, and $j = 1, ..., J$. Some entries of $Y$ are not observable. We use $z_{ij}$ to indicate the missing status of entry $Y_{ij}$, where $z_{ij} = 1$ indicates that $Y_{ij}$ is observed and $z_{ij} = 0$ otherwise. We let $Z = (z_{ij})_{N \times J}$ be the indicator matrix for data missingness. The main goal of 1-bit matrix completion is to estimate $E(Y_{ij} | z_{ij} = 0)$.

This problem is typically tackled under a probabilistic model (see e.g., Cai and Zhou, 2013; Davenport et al., 2014; Bhaskar and Javanmard, 2015), which assumes that $Y_{ij}$, $i = 1, ..., N$, $j = 1, ..., J$, are independent Bernoulli random variables, with success probability $\exp(m_{ij})/(1 + \exp(m_{ij}))$ or $\Phi(m_{ij})$, where $m_{ij}$ is a real-valued parameter and $\Phi$ is the cumulative distribution function of the standard normal distribution. It is further assumed that the matrix $M = (m_{ij})_{N \times J}$ is either exactly low-rank or approximately low-rank, where the approximate low-rankness is measured by the nuclear norm of $M$. Finally, a random sampling scheme is typically assumed for $z_{ij}$. For example, Davenport et al. (2014) considered a uniform sampling scheme where $z_{ij}$ are independent and identically distributed Bernoulli random variables, and Cai and Zhou (2013) considered a non-uniform sampling scheme. Under such a random sampling scheme, $Z$ and $Y$ are assumed to be independent and thus data missingness is ignorable.

It is of interest to draw statistical inference on linear forms of $M$, including the inference of individual entries of $M$. This is a challenging problem under the above general setting for 1-bit matrix completion, largely due to the presence of a non-linear link function. In particular, the
existing results on the inference for matrix completion as established in Xia and Yuan (2021) and Chen et al. (2019a) are under a linear model that observes $m_{ij} + \epsilon_{ij}$ for the non-missing entries, where $\epsilon_{ij}$ are mean-zero independent errors. Their analyses cannot be directly applied to non-linear models.

As the first inference work of 1-bit matrix completion with non-linear models, we start with a basic setting in which we assume the success probability takes a logistic form of $M$ and each $m_{ij}$ depends on a row effect and a column effect only. Asymptotic normality results are then established for the inference of $M$. Specifically, this model assumes that

1. given $M$, $Y_{ij}$, $i = 1, ..., N$, $j = 1, ..., J$, are independent Bernoulli random variables whose distributions do not depend on the missing indicators $Z$,

2. the success probability for $Y_{ij}$ is assumed to be $\exp(m_{ij})/(1+\exp(m_{ij}))$ that follows a logistic link,

3. $M$ has the model parameterization that $m_{ij} = \theta_i - \beta_j$.

In the rest, $\theta_i$ and $\beta_j$ will be referred to as the row and column parameters, respectively. This parameterization allows the success probability of each entry to depend on both a row effect and a column effect. We now introduce two real-world applications and discuss the interpretations of the row and column parameters in these applications.

**Example 1.** The introduced model is also referred to as the Rasch model, one of the most popular item response theory models (Embretson and Reise, 2013) to model item-level response data in educational testing and psychological measurement. In educational testing, each row of the data matrix represents an examinee and each column represents an item (i.e., exam question). Each binary entry $Y_{ij}$ records whether examinee $i$ correctly answers item $j$. The row parameter $\theta_i$ is interpreted as the ability of examinee $i$, which is an individual-specific latent factor, and the column parameter $\beta_j$ is interpreted as the difficulty of item $j$, as the probability of correctly answering an item increases with one’s ability $\theta_i$ and decreases with the difficulty level $\beta_j$ of the item.

In Section 5, we apply the considered model to link two forms of an educational test, an important practical issue in educational assessment (Kolen and Brennan, 2014). That is, consider two groups
of examinees taking two different forms of an educational test, where the two forms share some common items but not all, resulting in missingness of the data matrix. As the two test forms may have different difficulty levels, it is usually not fair to directly compare the total scores of two students who take different forms. The proposed method allows us to compare examinees’ performance as if they had taken the same test form and to also quantify the estimation uncertainty.

Example 2. Consider senators’ roll call voting records in the United States senate, and in this application, each row of the data matrix corresponds to a senator and each column corresponds to a bill voted in the senate. Each binary response $Y_{ij}$ records whether the senator voted for or against the bill. It has been well recognized in the political science literature (Poole et al., 1991; Poole and Rosenthal, 1991) that senate voting behavior is essentially unidimensional, though slightly different latent variable models are used in that literature. That is, it is believed that senators’ voting behavior is driven by a unidimensional latent factor, often interpreted as the conservative-liberal political ideology. Moreover, it is a consensus that the Republican senators tend to lie on the conservative side of the factor and the Democratic senators tend to lie on the liberal side, though there are sometimes a very small number of exceptions. To apply the our method to senators’ roll call voting records, we pre-process the data as follows. If bill $j$ is more supported by the Republican party than the Democratic party and senator $i$ voted for the bill, then we let $Y_{ij} = 1$. If bill $j$ is more supported by the Democratic party and senator $i$ voted against the bill, we let $Y_{ij} = 1$. Otherwise, $Y_{ij} = 0$. More details about this data pre-processing can be found in Section 5. Under the considered model, the row parameter may be interpreted as the conservativeness score of senator $i$. That is, the higher the conservativeness score of a senator, the higher chance for him/her to support a bill favored by the Republican party and to vote against a bill favored by the Democratic party. The column parameter characterises the bill effect.

In Section 5, we apply the model to link the roll call voting records from multiple years, where different senators have different terms in the senate, resulting in missingness of the data matrix. The model allows us to compare senators in terms of their conservative-liberal political ideology, even if they have not served in the senate at the same time.

As mentioned previously, the considered nonlinear factor model can be viewed as a Bradley-Terry model (Bradley and Terry, 1952) for directed graphs that is commonly used for modeling
pairwise comparisons. In Remark 1 below, we discuss this connection and explain the reason why the existing result such as Han et al. (2020) does not apply to the current setting.

**Remark 1.** Data $Y$ under our model setting can be viewed as a bipartite graph with $N + J$ nodes. Its adjacency matrix takes the form

$$
\begin{pmatrix}
NA_{N,N} & Y \\
(1_{N,J} - Y)^T & NA_{J,J}
\end{pmatrix},
$$

where $NA_{N,N}$ and $NA_{J,J}$ are two matrices whose entries are missing and $1_{N,J}$ is a matrix with all entries being 1. We let the value of $1 - Y_{ij}$ be missing if $Y_{ij}$ is missing (i.e., $z_{ij} = 0$). Such a directed graph can be modeled by the Bradley-Terry model; see Bradley and Terry (1952). In Han et al. (2020), asymptotic normality results are established for $n$-by-$n$ adjacency matrices that follow the Bradley-Terry model when the graph size $n$ grows to infinity. However, Han et al. (2020) only consider a uniformly missing setting. That is, the probability that the edges between two nodes are missing is assumed to be the same for all pairs of nodes. This assumption is not satisfied for the adjacency matrix (1), due to the two missing matrices on the diagonal. In fact, the asymptotic analysis under the current setting is more involved, due to the need of simultaneously considering two indices $N$ and $J$ and the increased complexity in approximating the asymptotic variance of model parameters.

Given data $\{Y_{ij} : z_{ij} = 1, i = 1, ..., N, j = 1, ..., J\}$, the log-likelihood function for parameters $\theta = (\theta_1, ..., \theta_N)^T$ and $\beta = (\beta_1, ..., \beta_J)^T$ takes the form

$$
l(\theta, \beta) = \sum_{i,j : z_{ij} = 1} [Y_{ij}(\theta_i - \beta_j) - \log(1 + \exp(\theta_i - \beta_j))].
$$

The identifiability of parameters $\theta$ and $\beta$ is subject to a location shift. That is, the distribution of data remains unchanged, if we add a common constant to all the $\theta_i$ and $\beta_j$, as the likelihood function in (2) only depends on the all differences $\theta_i - \beta_j$. To avoid ambiguity, we require $\sum_{i=1}^{N} \theta_i = 0$ in the rest. We point out that this requirement does not play a role when we draw inference about any linear form of $M$, as the location shift of $\theta$ and $\beta$ does not affect the value of $M$. We estimate
\(\theta\) and \(\beta\) by the maximum likelihood estimator

\[
(\hat{\theta}, \hat{\beta}) = \arg \max_{\theta, \beta} l(\theta, \beta), \text{s.t., } \sum_{i=1}^{N} \theta_i = 0.
\]

The maximum likelihood estimator of \(\theta\) and \(\beta\) further leads to the maximum likelihood estimator of \(M\), \(\hat{m}_{ij} = \hat{\theta}_i - \hat{\beta}_j\). It is easy to see that (3) is a convex optimization problem. Thanks to the low-rank structure of \(M\), this problem can be efficiently solved by performing alternating maximization, as often used for estimating low-rank matrices (Chen et al., 2019b, 2020b; Udell et al., 2016). Such an algorithm is implemented for the numerical experiments, whose details are provided in Algorithm 1 below.

**Algorithm 1: Alternating Gradient Ascent Algorithm**

**Input:** Partially observed data matrix \(Y\), learning rate \(\gamma\), tolerance threshold \(tol\).

**Output:** Estimates \(\hat{\theta}, \hat{\beta}\).

Initialize \(\theta^{(0)} = \{\theta_i^{(0)} : i = 1, \ldots, N\}, \beta^{(0)} = \{\beta_j^{(0)} : j = 1, \ldots, J\}\)

with \(\theta_i^{(0)}, \beta_j^{(0)} \sim \text{Uniform}(c, c)\), \(i = 1, \ldots, N, j = 1, \ldots, J\),

\[JML^{(0)} = 0, JML^{(1)} = \sum_{i,j: z_{ij}=1} (y_{ij} \{\theta_i^{(0)} - \beta_j^{(0)}\} - \log [1 + \exp \{\theta_i^{(0)} - \beta_j^{(0)}\}]);\]

while \(JML^{(1)} - JML^{(0)} > tol\) do

\(JML^{(0)} = JML^{(1)};\)

for \(i = 1, \ldots, N\) do

\(\theta_i^{(1)} \leftarrow \theta_i^{(0)} + \gamma \left( \sum_{j: z_{ij}=1} y_{ij} - \{e^{\theta_i^{(0)} - \beta_j^{(0)}}\} / \{1 + e^{\theta_i^{(0)} - \beta_j^{(0)}}\} \right);\)

end

for \(j = 1, \ldots, J\) do

\(\beta_j^{(1)} \leftarrow \beta_j^{(0)} + \gamma \left( \sum_{i: z_{ij}=1} y_{ij} + \{e^{\theta_i^{(1)} - \beta_j^{(0)}}\} / \{1 + e^{\theta_i^{(1)} - \beta_j^{(0)}}\} \right);\)

end

\[
\theta^{(1)} = \theta^{(1)} - N^{-1} \sum_{i=1}^{N} \theta_i^{(1)};
\]

\[
\beta^{(1)} = \beta^{(1)} - N^{-1} \sum_{i=1}^{N} \beta_i^{(1)};
\]

\[
JML^{(1)} = \sum_{i,j: z_{ij}=1} (y_{ij} \{\theta_i^{(1)} - \beta_j^{(1)}\} - \log [1 + \exp \{\theta_i^{(1)} - \beta_j^{(1)}\}]);\]

\(\theta^{(0)} = \theta^{(1)};\)

\(\beta^{(0)} = \beta^{(1)};\)

end
3 Statistical Inference

In this section, we consider the statistical inference of any linear form of $M$. Specifically, we use $g : \mathbb{R}^{N \times J} \rightarrow \mathbb{R}$ to denote a linear function of $M$ that takes the form

$$g(M) = \sum_{i=1}^{N} \sum_{j=1}^{J} w_{ij} m_{ij}, \quad (4)$$

where the weights $w_{ij}$ are pre-specified. It is straightforward that a point estimate of $g(M)$ is given by $g(\hat{M}) = \sum_{i=1}^{N} \sum_{j=1}^{J} w_{ij} \hat{m}_{ij}$. Our goal is to establish the asymptotic normality for $g(\hat{M})$, based on which we can test hypothesis about $g(M)$ or construct confidence intervals. We provide two examples of $g(M)$ that may be of interest in practice.

**Example 3.** Consider $g(M) = m_{ij}$ for entry $(i,j)$ that is not observed, i.e., $z_{ij} = 0$. The asymptotic normality of $\hat{m}_{ij}$ allows us to quantify the uncertainty in our prediction $\exp(\hat{m}_{ij})/(1 + \exp(\hat{m}_{ij}))$ of the unobserved entry.

**Example 4.** Consider $g(M) = \sum_{j=1}^{J}(m_{ij} - m_{i'j})/J = \theta_i - \theta_{i'}$, that is of interest in both educational testing and ranking. If we interpret the model as the Rasch model in educational testing, then $\theta_i$ can be regarded as examinee $i$'s ability level. Examinee $i$ is more likely to answer any question correctly than examinee $i'$ if $\theta_i > \theta_{i'}$, and vice versa. Therefore, even when two examinees do not answer the same test form, the statistical inference of this quantity will allow us to compare their performance and further quantify the uncertainty in this comparison. On the other hand, if we draw connections to Bradley-Terry model in ranking, then $\theta_i$ can be interpreted as subject $i$'s ranking criteria. The statistical inference on $(\theta_i - \theta_{i'})$ for any combination of $i,i'$ would allow us to quantify the uncertainty in the rankings of all $N$ subjects.

We first establish the existence and consistency for $M$, $\theta$, and $\beta$. We denote $J_* = \min \left\{ \sum_{j=1}^{J} z_{ij} : i = 1, \ldots, N \right\}$ and $J^* = \max \left\{ \sum_{j=1}^{J} z_{ij} : i = 1, \ldots, N \right\}$ as the minimum and maximum numbers of observed entries per row, respectively. Similarly, we
denote
\[ N_\ast = \min \left\{ \sum_{i=1}^{N} z_{ij} : j = 1, \ldots, J \right\} \quad \text{and} \quad N^\ast = \max \left\{ \sum_{i=1}^{N} z_{ij} : j = 1, \ldots, J \right\} \]
as the minimum and maximum numbers of observed entries per column, respectively. Let \( \|x\|_\infty = \max\{|x_i| : i = 1, \ldots, n\} \) be the infinity norm of a vector \( x = (x_1, \ldots, x_n)^T \). Let \( \theta^\ast, \beta^\ast \) and \( M^\ast \) be the true values of \( \theta, \beta \) and \( M \), respectively. Without loss of generality, we assume \( N \geq J \). For simplicity, we also assume \( N_\ast > J^\ast \) and \( N^\ast > J^\ast \). The following conditions are required.

**Condition 1.** As \( N \) and \( J \) grow to infinity, the following are satisfied:

(a) There exists a constant \( k > 0 \), such that \( N_\ast \geq kN^{2/3} \) and \( J_\ast \geq kJ^{2/3} \);

(b) \( (J_\ast)^{-1}(\log N) \) converge to 0;

(c) There exist positive constants \( k_1 \) and \( k_2 \) such that \( k_1 J_\ast \leq J^\ast \leq k_2 J_\ast \).

**Condition 2.** There exists a constant \( c < \infty \) such that \( \|\theta^\ast\|_\infty < c \) and \( \|\beta^\ast\|_\infty < c \).

**Condition 3.** For any \((i,j)\), there exists \( 1 \leq i_1, i_2, \ldots, i_k \leq N \) and \( 1 \leq j_1, j_2, \ldots, j_k \leq J \) such that \( z_{ij_1} = z_{i_1j_1} = z_{i_1j_2} = z_{i_2j_2} = \ldots = z_{i_kj_k} = z_{i_kj} = 1 \).

We provide some discussions on Conditions 1 and 3. Condition 1(a) requires the number of observations for each parameter to grow to infinity at a suitable rate. Under this requirement, the proportion of observable entries is allowed to decay to zero at the rate \( (NJ)^{-\frac{2}{3}} \). Condition 1(b) is a very mild technical condition that requires \( J_\ast \) to grow faster than \( \log(N) \). Condition 1(c) requires that \( J_\ast \) and \( J^\ast \) are of the same order. This assumption essentially requires a balanced missing data pattern that has a similar spirit as the random sampling regimes for missingness adopted in Cai and Zhou (2013) and Davenport et al. (2014). Condition 3 is necessary and sufficient for the identifiability of \( \theta \) and \( \beta \); see Proposition 1 for a formal statement.

**Proposition 1.** If Condition 3 holds, then \( \theta \) and \( \beta \) are uniquely determined by equations \( \sum_{i=1}^{N} \theta_i = 0 \) and \( \theta_i - \beta_j = m_{ij}, i = 1, \ldots, N, j = 1, \ldots, J \), for which \( z_{ij} = 1 \).

If Condition 3 does not hold, then there exists \( (\tilde{\theta}, \tilde{\beta}) \neq (\theta, \beta) \), such that \( \sum_{i=1}^{N} \tilde{\theta}_i = 0, \sum_{i=1}^{N} \theta_i = 0, \) and \( \theta_i - \beta_j = \tilde{\theta}_i - \tilde{\beta}_j, i = 1, \ldots, N, j = 1, \ldots, J, z_{ij} = 1 \).
Theorem 1 below guarantees the existence and consistency of the maximum likelihood estimator, when both \(N\) and \(J\) grow to infinity.

**Theorem 1.** Assume Conditions 1, 2 and 3 hold. Then, as \(N, J\) grow to infinity, maximum likelihood estimator \((\hat{\theta}, \hat{\beta})\) exists, with probability tending to 1. Furthermore, as \(N\) and \(J\) grow to infinity, we have

\[
\|\hat{\theta} - \theta^*\|_\infty = O_p\{(\log N)^{\frac{1}{2}}J_*^{-\frac{1}{2}}\}, \quad \|\hat{\beta} - \beta^*\|_\infty = O_p\{(\log J)^{\frac{1}{2}}N_*^{-\frac{1}{2}}\},
\]

and

\[
\max_{i,j} |\hat{m}_{ij} - m_{ij}^*| = O_p\{(\log J)^{\frac{1}{2}}N_*^{-\frac{1}{2}} + (\log N)^{\frac{1}{2}}J_*^{-\frac{1}{2}}\}.
\]

To state the asymptotic normality result for \(g(\hat{M})\), we reexpress

\[g(M) = w_g^T \theta + \tilde{w}_g^T \beta,\]

where \(w_g = (w_{g1}, \ldots, w_{gN})^T\) and \(\tilde{w}_g = (\tilde{w}_{g1}, \ldots, \tilde{w}_{gJ})\). Note that this expression always exists, by letting \(w_{gi} = \sum_{j=1}^J w_{ij}\) and \(\tilde{w}_{gj} = -\sum_{i=1}^N w_{ij}\). We introduce some notations. Let \(\sigma^2_{ij} = \text{var}(Y_{ij}) = \exp(\theta_i^* - \beta_j^*)/\{1 + \exp(\theta_i^* - \beta_j^*)\}^2\), \(\sigma^2_{i+} = \sum_{j=1}^J z_{ij}\sigma^2_{ij}\), and \(\sigma^2_{+j} = \sum_{i=1}^N z_{ij}\sigma^2_{ij}\). Further denote \(\hat{\sigma}^2_{ij} = \exp(\hat{\theta}_i - \hat{\beta}_j)/\{1 + \exp(\hat{\theta}_i - \hat{\beta}_j)\}^2\), \(\hat{\sigma}^2_{i+} = \sum_{j=1}^J z_{ij}\hat{\sigma}^2_{ij}\), and \(\hat{\sigma}^2_{+j} = \sum_{i=1}^N z_{ij}\hat{\sigma}^2_{ij}\) to be the corresponding plug-in estimates. We use \(\|\cdot\|_1\) to denote the \(L_1\) norm of a vector. The result is summarized in Theorem 2 below.

**Theorem 2.** Assume Conditions 1, 2 and 3 hold and \(J_*^{-2}N_* (\log N)^2 \to 0\) as \(N \to \infty\). Consider a linear function \(g(M) = w_g^T \theta + \tilde{w}_g^T \beta\) with \(g(M) \neq 0\). Further suppose that there exists a constant \(C > 0\) such that \(\|w_g\|_1 < C\) and \(\|\tilde{w}_g\|_1 < C\). Then

\[\hat{\sigma}(g)^{-1}\{g(\hat{M}) - g(M^*)\} \to N(0, 1)\text{ in distribution,}\]

where \(\hat{\sigma}^2(g) = \sum_{i=1}^N w_{gi}^2 (\sigma^2_{i+})^{-1} + \sum_{j=1}^J \tilde{w}_{gj}^2 (\sigma^2_{+j})^{-1}\).

Moreover, \(\hat{\sigma}(g)\) can be replaced by its plug-in estimator, i.e.,

\[\hat{\sigma}(g)^{-1}\{g(\hat{M}) - g(M^*)\} \to N(0, 1)\text{ in distribution,}\]

(5)
where \( \hat{\sigma}^2(g) = \sum_{i=1}^{N} w_{gi}^2 (\hat{\sigma}_{i+}^2)^{-1} + \sum_{j=1}^{J} \tilde{w}_{gj}^2 (\hat{\sigma}_{+j}^2)^{-1}. \)

We now discuss the implications of Theorem 2. For each \( \theta_i \), \( \text{var}(\hat{\theta}_i) = (\sigma_{i+}^2)^{-1} \{1 + o(1)\} \). It is worth noting that by the classical theory for maximum likelihood estimation, \( (\sigma_{i+}^2)^{-1} \) is the Cramer-Rao lower bound for the estimation of \( \theta_i \), when the column parameters \( \beta \) are known. Thus, the result of Theorem 2 implies that \( \hat{\theta}_i \) is an asymptotically optimal estimator for \( \theta_i \). Similarly, for each \( \beta_j \), \( \text{var}(\hat{\beta}_j) = (\sigma_{+j}^2)^{-1} \{1 + o(1)\} \), which also achieves the Cramer-Rao lower bound asymptotically, when the row parameters \( \theta \) are known. Moreover, \( \text{var}(\hat{m}_{ij}) = \text{var}(\hat{\theta}_i - \hat{\beta}_j) \equiv \{(\sigma_{i+}^2)^{-1} + (\sigma_{+j}^2)^{-1}\} \{1 + o(1)\} \). We end this section with a remark.

**Remark 2.** The derived asymptotic theory is different from that for non-linear regression models of increasing dimensions that has been studied in Portnoy (1988), He and Shao (2000) and Wang (2011). To achieve asymptotic normality under the setting of these works, one at least requires the number of observations to grow faster than the square of the number of parameters. Under the setting of the current work, the model has \( N + J - 1 \) free parameters, while the number of observed entries is allowed to grow as slow as \( O((NJ)^{3/4}) \), which is much slower than \( (N + J - 1)^2 \). Even when there is no missing entries, the number of observed entries is \( NJ \) which does not grow as fast as \( (N + J - 1)^2 \).

## 4 Simulation Study

We study the finite-sample performance of the likelihood-based estimator. We consider two settings: (1) \( N = 5000 \) and \( J = 200 \), and (2) \( N = 10000 \) and \( J = 400 \). Missing data are generated under a block-wise design. That is, we split the rows into five equal-sized clusters and the columns into four equal-sized clusters. We let each row cluster correspond to the columns from a distinct combination of two column clusters. Rows from the same cluster have the same missing pattern. Specifically, their entries are observable and only observable, on the columns that this row cluster correspond to. This missing data pattern can be illustrated by a five-by-four block-wise matrix \( \{(1,0,0,1,0)^T, (1,1,0,0,1)^T, (0,1,1,1,0)^T, (0,0,1,0,1)^T\} \), where 1 and 0 represent a sub-matrix with \( z_{ij} = 1 \) and 0, respectively. An illustration of the missing pattern \( Z \) is illustrated in Figure 1. Under the first setting, \( N^* = 2000, N^* = 3000, \) and \( J^* = J^* = 100 \). Under the second
setting, \(N_\ast = 4000, N^\ast = 6000\), and \(J_\ast = J^\ast = 200\). For each setting, \(\theta\) is simulated from a uniform distribution over the space \(\{x = (x_1, \ldots, x_N)^T : \sum_{i=1}^N x_i = 0, -2 \leq x_i \leq 2\}\), and \(\beta\) is obtained by simulating \(\beta_j\) independently from the uniform distribution over the interval \([-2, 2]\). For each setting, 2000 independent datasets are generated from the considered model.

![Figure 1: A heat map of \(Z\). The black and white regions correspond to \(z_{ij} = 1\) and \(0\), respectively.](image)

Under setting (1), the mean squared estimation errors for \(M\), \(\theta\) and \(\beta\) are 0.067, 0.064 and 0.0028, respectively, across all relevant entries and all 2000 independent samples. Under setting (2), these values read 0.033, 0.031 and 0.0013, respectively. Unsurprisingly, increasing sample sizes can improve estimation accuracy.

We then examine the variance approximation in Theorem 2. We compare \(\hat{\sigma}^2(g)\), \(\tilde{\sigma}^2(g)\) and \(s^2(g)\), where \(s^2(g)\) denotes the sample variance of \(g(\hat{M})\) that is calculated based on the 2000 simulations. As \(\sigma^2(g)\) varies across the datasets, we calculate \(\hat{\sigma}^2(g)\) as the average of \(\tilde{\sigma}^2(g)\) over 2000 simulated datasets. We consider functions \(g(M) = m_{ij}, \theta_i, \beta_j, i = 1, \ldots, N, j = 1, \ldots, J\). The results are given in Figure 2, where panels (a)-(c) show the scatter plots of \(s^2(g)\) against \(\hat{\sigma}^2(g)\) and panels (d)-(f) show those of \(s^2(g)\) against \(\tilde{\sigma}^2(g)\). These plots suggest that \(\hat{\sigma}^2(g), \tilde{\sigma}^2(g),\) and \(s^2(g)\) are close to each other, for the specific forms of \(g\) that are examined.

To validate asymptotic normality, we compare the empirical densities of the 2000 sample estimates of \(m_{11}, \theta_1\) and \(\beta_1\) against their respective theoretical normal density curves in Figure 3 for illustration. We can observe from Figure 3 that the empirical distributions of the estimates agree well with their corresponding theoretical distributions.

Furthermore, for each \(m_{ij}, \theta_i,\) and \(\beta_j\), we construct its 95% Wald interval based on (5), for which the empirical coverage based on 2000 independent replications is computed. This result is shown in Figure 4, where the two panels correspond to the two simulation settings, respectively. In each panel, the three box-plots show the empirical coverage probabilities for entries of \(M\), \(\theta\), and \(\beta\), respectively. As we can see, all these empirical coverage probabilities are close to the nominal
Figure 2: Panels (a)-(c) plot $s^2(g)$ against $\hat{\sigma}^2(g)$ for $g(M) = m_{ij}$, $\theta_i$, and $\beta_j$, respectively. Panels (d)-(f) plot $s^2(g)$ against $\hat{\sigma}^2(g)$ for $g(M) = m_{ij}$, $\theta_i$ and $\beta_j$, respectively. Each panel shows 100 randomly sampled $m_{ij}$, $\theta_i$, or $\beta_j$ under each setting. The line $y = x$ is given as a reference.

5 Real-data Applications

In what follows, we consider two real-data applications.

5.1 Application to Educational Testing

We first apply the proposed method to link two forms of an educational test that share common items. The dataset is a benchmark dataset for studying linking methods for educational testing (González and Wiberg, 2017). It contains binary responses from two forms of a college admission test. Each form has 120 items and is answered by 2000 examinees. There are 40 common items shared by the two test forms. There is no missing data within each test. Thus, $N = 4000$, $J = 200$, and 40% of the data entries are missing. We apply the proposed method to this dataset. Making use of Theorem 2, 95% confidence intervals are obtained for both the row (i.e., person) parameters and
the column (i.e., item) parameters. The results allow us to compare students who took different test forms, as well as non-common items from the two forms. For illustration, we randomly choose 100 row parameters and 100 column parameters and show their 95% confidence intervals in Figure 5. Such uncertainty quantification can be vital for colleges when making admission decisions.

5.2 Application to Senate Voting

We now apply the proposed method to the United States senate roll call voting data. Data from the 111th through the 113th congress that include the voting records from January 11, 2009 to December 16, 2014. Quite a few senators did not serve for the entire period.

To apply our method to senators’ roll call voting records with $\theta_i$ being interpreted as the conservativeness score of senator $i$, we pre-process the data as follows. First, five senators who did not serve for more than half a year during the period are removed from the dataset, including Edward M. Kennedy, Joe Biden, Hilary Clinton, Julia Salazar and Carte Goodwin. Second, 191
Figure 4: Panels (a) and (b) show the empirical coverage rates for the 95% Wald intervals under settings (1) and (2), respectively.

Figure 5: (a) 95% confidence intervals of 100 row parameters, with 50 randomly selected from each group. (b) 95% confidence intervals of the 100 column parameters, with 40 each randomly chosen from group 1 and group 2 and 20 randomly selected from anchor items (i.e., common items).

bills are removed, as all the observed votes to each of these bills are the same and consequently their maximum likelihood estimates do not exist. After these two steps, the resulting dataset contains $N = 139$ senators and $J = 1648$ bills. Finally, for bill $j$ that has a higher percentage support within the Republican party than that within the Democratic party, we let $Y_{ij} = 1$ if senator $i$ voted for the bill and $Y_{ij} = 0$ if senator $i$ voted against it. For bill $j$ that has a higher percentage support within the Democratic party than that within the Republican party, we let $Y_{ij} = 1$ if senator $i$ voted against the bill and $Y_{ij} = 0$ if he/she voted for it. The value of $Y_{ij}$ is missing, if the senator chose not to vote or he/she was not in the senate when this bill was voted. For the final data being analyzed, the proportion of missing entries is 26.1% and the connectedness Condition 3 is satisfied. The missingness pattern of the dataset is given in Figure 6. Note that in this example, $N < J$. However, our asymptotic results are still applicable if we simply switch the roles of $N$ and $J$ in the required conditions.

Our asymptotic results allow us to compare senators’ ideological position, even if they did not
serve in the senate at the same time. For example, Judd Gregg served in the senate between January 3, 1993 and January 3, 2011, while Marco Rubio started his first term as a senator since January 3, 2011. In our model, Judd Gregg ($\theta_i$) and Marco Rubio ($\theta_k$) have estimated conservativeness scores of 2.59 and 4.25, respectively. Applying our asymptotic results, we have $\hat{\theta}_i - \hat{\theta}_k = -1.66$ and its standard error is 0.169. If we test $H_0 : \theta_i = \theta_k$ against $H_1 : \theta_i \neq \theta_k$, we obtain an extremely small p-value of $9.0 \times 10^{-23}$. Therefore, we conclude that senator Marco Rubio is significantly more conservative than senator Judd Gregg.

In addition, we present in Tables 1 and 2 the ten senators with the largest row parameter estimates, and the ten senators with the smallest row parameter estimates. These results align well with the public perceptions about these senators. For example, Jim Demint, who is ranked the most conservative senator in this dataset by our method, was also identified by Salon as one of the most conservative members of the senate (Kornacki, 2011). Our method ranks Mike Lee the second, though his conservativeness score is not significantly different from that of Demint. In fact, in 2017, the New York Times used the NOMINATE system (Poole and Rosenthal, 2001) to arrange Republican senators by ideology and ranked Lee as the most conservative member of the Senate (Parlapiano et al., 2017). For another example, Brian Schatz who is ranked to be the most liberal senator by our method, is well-known as a liberal Democrat. During his time in the senate, he voted with the Democratic party on most issues.

Finally, the 95% confidence intervals for all the row parameters are shown in Figure 7 and a full list of rankings for all the 139 senators is given in the Appendix section, where the corresponding row parameter estimates and their standard errors are also presented.

Figure 6: A heat map of $Z$. The black and white regions correspond to $z_{ij} = 1$ and 0, respectively.
Table 1: Ranking of the top 10 most conservative senators predicted by the model. Rep and Dem represent the Republican party and the Democratic party, respectively.

| Rank | Senator (party) | State            | Conservativeness Score (s.e.(θ)) |
|------|-----------------|------------------|----------------------------------|
| 1    | Jim DeMint (Rep)| South Carolina   | 5.87 (0.157)                     |
| 2    | Mike Lee (Rep)  | Utah             | 5.73 (0.138)                     |
| 3    | Ted Cruz (Rep)  | Texas            | 5.65 (0.195)                     |
| 4    | Tom Coburn (Rep)| Oklahoma         | 5.25 (0.114)                     |
| 5    | Rand Paul (Rep)| Kentucky         | 5.24 (0.129)                     |
| 6    | Tim Scott (Rep)| South Carolina   | 5.17 (0.176)                     |
| 7    | Jim Bunning (Rep)| Kentucky     | 4.92 (0.204)                     |
| 8    | Ron Johnson (Rep)| Wisconsin   | 4.84 (0.119)                     |
| 9    | James Risch (Rep)| Idaho        | 4.81 (0.102)                     |
| 10   | Jim Inhofe (Rep)| Oklahoma        | 4.69 (0.103)                     |

Table 2: Ranking of the top 10 most liberal senators predicted by the model. Rep and Dem represent the Republican party and the Democratic party, respectively.

| Rank | Senator (party) | State  | Conservativeness Score (s.e.(θ)) |
|------|-----------------|--------|----------------------------------|
| 1    | Brian Schatz (Dem)| Hawaii | -4.74 (0.468)                   |
| 2    | Roland Burris (Dem)| Illinois | -4.43 (0.297)                   |
| 3    | Mazie Hirono (Dem)| Hawaii | -4.17 (0.383)                   |
| 4    | Cory Booker (Dem)| New Jersey | -4.14 (0.572)                   |
| 5    | Tammy Baldwin (Dem)| Wisconsin | -3.90 (0.352)                   |
| 6    | Sherrod Brown (Dem)| Ohio  | -3.89 (0.168)                   |
| 7    | Tom Udall (Dem) | New Mexico | -3.85 (0.165)                   |
| 8    | Dick Durbin (Dem)| Illinois | -3.83 (0.164)                   |
| 9    | Ben Cardin (Dem) | Maryland | -3.82 (0.163)                   |
| 10   | Sheldon Whitehouse (Dem)| Rhode Island | -3.74 (0.163)                   |

6 Discussions

This note considers the statistical inference for 1-bit matrix completion under a unidimensional nonlinear factor model, the Rasch model. Asymptotic normality results are established. Our results suggest that the maximum likelihood estimator is statistically efficient, even though the number of parameters diverges. Our simulation study shows that the developed asymptotic result provides a good approximation to finite sample data, and two real-data examples demonstrate its usefulness in the areas of educational testing and political science.

The current results can be easily extended to matrix completion problems with quantized measurement that has a similar natural exponential family form. Admittedly, the model considered may be oversimple for complex application problems, for example, certain collaborative filtering
problems for which the rank of the underlying matrix $M$ may be higher than considered here and the underlying latent factors may be multidimensional. The extension of the current results to more flexible models is left for future investigation. As the first inference result for 1-bit matrix completion, we believe the current results will shed light on the statistical inference for more general matrix completion problems.

References

Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society: Series B (Methodological)*, 32(2):283–301.

Bahadur, R. R. (1971). Some limit theorems in statistics. SIAM.

Berk, R. H. (1972). Consistency and asymptotic normality of MLE’s for exponential models. *The Annals of Mathematical Statistics*, 43:193–204.

Bhaskar, S. A. (2016). Probabilistic low-rank matrix completion from quantized measurements. *The Journal of Machine Learning Research*, 17(1):2131–2164.

Bhaskar, S. A. and Javanmard, A. (2015). 1-bit matrix completion under exact low-rank constraint. In *2015 49th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6. IEEE.
Biswas, P., Lian, T., Wang, T., and Ye, Y. (2006). Semidefinite programming based algorithms for sensor network localization. *ACM Transactions on Sensor Networks (TOSN)*, 2(2):188–220.

Bradley, R. A. and Terry, M. E. (1952). Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345.

Cai, T. and Zhou, W. (2013). A max-norm constrained minimization approach to 1-bit matrix completion. *The Journal of Machine Learning Research*, 14:3619–3647.

Candès, E. J. and Recht, B. (2009). Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 9(6):717–772.

Candès, E. J. and Tao, T. (2010). The power of convex relaxation: Near-optimal matrix completion. *IEEE Transactions on Information Theory*, 56(5):2053–2080.

Chatterjee, S., Diaconis, P., Sly, A., et al. (2011). Random graphs with a given degree sequence. *Annals of Applied Probability*, 21(4):1400–1435.

Chen, Y., Chi, Y., Fan, J., Ma, C., and Yan, Y. (2020a). Noisy matrix completion: Understanding statistical guarantees for convex relaxation via nonconvex optimization. *SIAM Journal on Optimization*, 30(4):3098–3121.

Chen, Y., Fan, J., Ma, C., and Yan, Y. (2019a). Inference and uncertainty quantification for noisy matrix completion. *Proceedings of the National Academy of Sciences*, 116(46):22931–22937.

Chen, Y., Li, X., and Zhang, S. (2019b). Joint maximum likelihood estimation for high-dimensional exploratory item factor analysis. *Psychometrika*, 84(1):124–146.

Chen, Y., Li, X., and Zhang, S. (2020b). Structured latent factor analysis for large-scale data: Identifiability, estimability, and their implications. *Journal of the American Statistical Association*, 115:1756–1770.

Davenport, M. A., Plan, Y., Van Den Berg, E., and Wootters, M. (2014). 1-bit matrix completion. *Information and Inference: A Journal of the IMA*, 3:189–223.

Embretson, S. E. and Reise, S. P. (2013). *Item Response Theory*. Psychology Press.
Goldberg, D., Nichols, D., Oki, B. M., and Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM, 35*(12):61–70.

González, J. and Wiberg, M. (2017). *Applying test equating methods*. Springer, New York.

Gutman, R., DeDe, G., Caplan, D., and Liu, J. S. (2011). Rasch model and its extensions for analysis of aphasic deficits in syntactic comprehension. *Journal of the American Statistical Association, 106*(496):1304–1316.

Haberman, S. J. (1977). Maximum likelihood estimates in exponential response models. *The Annals of Statistics, 5*:815–841.

Hagquist, C. and Andrich, D. (2017). Recent advances in analysis of differential item functioning in health research using the Rasch model. *Health and Quality of Life Outcomes, 15*(1):1–8.

Han, R., Ye, R., Tan, C., Chen, K., et al. (2020). Asymptotic theory of sparse bradley–terry model. *Annals of Applied Probability, 30*:2491–2515.

He, X. and Shao, Q. (2000). On parameters of increasing dimensions. *Journal of Multivariate Analysis, 73*(1):120–135.

Kantorovich, L. and Akilov, G. (1964). *Functional analysis in normed spaces (translated by D.G. Brown)*. Pergamon Presss, Oxford.

Keshavan, R. H., Montanari, A., and Oh, S. (2010). Matrix completion from noisy entries. *The Journal of Machine Learning Research, 11*:2057–2078.

Kidwell, P., Lebanon, G., and Collins-Thompson, K. (2011). Statistical estimation of word acquisition with application to readability prediction. *Journal of the American Statistical Association, 106*(493):21–30.

Klopp, O., Lafond, J., Moulines, E., and Salmon, J. (2015). Adaptive multinomial matrix completion. *Electronic Journal of Statistics, 9*(2):2950–2975.

Kolen, M. J. and Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices*. Springer Science & Business Media.
Koltchinskii, V., Lounici, K., and Tsybakov, A. B. (2011). Nuclear-norm penalization and optimal rates for noisy low-rank matrix completion. *The Annals of Statistics*, 39(5):2302–2329.

Kornacki, S. (2011). Why healthcare may not doom mitt romney after all. https://www.salon.com/2011/05/12/romney_healthcare/.

Lindsay, B., Clogg, C. C., and Grego, J. (1991). Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *Journal of the American Statistical Association*, 86(413):96–107.

Liu, Z. and Vandenberghe, L. (2010). Interior-point method for nuclear norm approximation with application to system identification. *SIAM Journal on Matrix Analysis and Applications*, 31(3):1235–1256.

Negahban, S. and Wainwright, M. J. (2012). Restricted strong convexity and weighted matrix completion: optimal bounds with noise. *The Journal of Machine Learning Research*, 13(1):1665–1697.

Parlapiano, A., Andrews, W., and Lee, J.and Shorey, R. (2017). How each senator voted on obamacare repeal proposals. https://www.nytimes.com/interactive/2017/07/25/us/politics/senate-votes-repeal-obamacare.html.

Poole, K. T. and Rosenthal, H. (1991). Patterns of congressional voting. *American Journal of Political Science*, 35:228–278.

Poole, K. T. and Rosenthal, H. (2001). Dnominate after 10 years: A comparative update to congress: a political-economic history of roll-call voting. *Legislative Studies Quarterly*, 26:5–29.

Poole, K. T., Rosenthal, H., and Koford, K. (1991). On dimensionalizing roll call votes in the us congress. *The American Political Science Review*, 85:955–976.

Portnoy, S. (1988). Asymptotic behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *The Annals of Statistics*, 16:356–366.

Rao, C. R. (1973). *Linear statistical inference and its applications*. Wiley New York.
Rasch, G. (1960). *Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests*. Nielsen & Lydiche, Oxford, England.

Rice, K. M. (2004). Equivalence between conditional and mixture approaches to the Rasch model and matched case-control studies, with applications. *Journal of the American Statistical Association*, 99(466):510–522.

Rinaldo, A., Petrović, S., and Fienberg, S. E. (2013). Maximum likelihood estimation in the β-model. *The Annals of Statistics*, 41:1085–1110.

Simons, G. and Yao, Y. (1999). Asymptotics when the number of parameters tends to infinity in the bradley-terry model for paired comparisons. *The Annals of Statistics*, 27:1041–1060.

Udell, M., Horn, C., Zadeh, R., and Boyd, S. (2016). Generalized low rank models. *Foundations and Trends® in Machine Learning*, 9:1–118.

van der Linden, W. J. and Hambleton, R. K. (2013). *Handbook of modern item response theory*. Springer, New York, NY.

Wang, L. (2011). GEE analysis of clustered binary data with diverging number of covariates. *The Annals of Statistics*, 39(1):389–417.

Xia, D. and Yuan, M. (2021). Statistical inferences of linear forms for noisy matrix completion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 83(1):58–77.
A Appendix

The appendix section contains the proofs of theorems and proposition in Section A.1, the proofs of the supporting lemmas in Section A.2, and a full list of rankings of senators considered in Section 5.2 of the main article according to their conservativeness scores in Section A.3

A.1 Proofs of Theorems and Proposition

Proof of Theorem 1. We start with defining some notations. Implicitly index $J$ with $N$ such that $J_N \to \infty$ as $N \to \infty$ for notation convenience. Note that this does not impose any rate requirement for $N$ and $J$. Let $\Omega_N = \{x = (x_{ij} : z_{ij} = 1, i = 1, ..., N, j = 1, ..., J) : x_{ij} = \theta_i - \beta_j, \theta_i, \beta_j \in \mathbb{R}, \sum_{i=1}^N \theta_i = 0\}$ be a vector space. Define on $\Omega_N$ a variance weighted inner product $\langle \cdot, \cdot \rangle_\sigma$ with $\langle x, y \rangle_\sigma = \sum_{i=1}^N \sum_{j \in S_J(i)} x_{ij} \sigma^2_{ij} y_{ij}$ for any $x, y \in \Omega_N$, where $S_J(i) = \{j = 1, ..., J : z_{ij} = 1\}$, $\sigma^2_{ij} = \exp(m_{ij}^*)/(1 + \exp(m_{ij}^*))^2$ and the subscript $\sigma$ means the inner product depends on $\sigma^2_{ij}, i = 1, ..., N, j = 1, ..., J, z_{ij} = 1$. Denote the associated norm as $\|\cdot\|_\sigma$ with $\|x\|_\sigma^2 = \sum_{i=1}^N \sum_{j \in S_J(i)} x_{ij}^2 \sigma^2_{ij}$ for $x \in \Omega_N$. Let $M_N = (m_{ij} : z_{ij} = 1, i = 1, ..., N, j = 1, ..., J, m_{ij} = \theta_i - \beta_j) \in \Omega_N$, $M^*_N = (m^*_{ij} : z_{ij} = 1, i = 1, ..., N, j = 1, ..., J, m^*_{ij} = \theta^*_i - \beta^*_j) \in \Omega_N$ and $\tilde{M}_N = (m_{ij} : z_{ij} = 1, i = 1, ..., N, j = 1, ..., J, \tilde{m}_{ij} = \tilde{\theta}_i - \tilde{\beta}_j) \in \Omega_N$. Note that as a result of Proposition 1, for any linear form $g$ of $M$, $g(M)$ can be re-expressed as a linear form of $x \in \Omega_N$, with $g(x) = \sum_{i=1}^N \sum_{j \in S_J(i)} w_{ij} x_{ij}$, where we denote $w_{ij} = w_{ij}(g)$, which depends on $g$, for notation simplicity. Let $\Omega_N^*$ consist of all linear forms $g$ on $\Omega_N$ such that $g(x) = 0$ if $x = 0$ and $x \in \Omega_N$. Without loss of generality, we will work with $g \in \Omega_N^*$ in the proofs. For any subset $A \subset \Omega_N^*$, define $\|\cdot\|_\sigma(A)$ to be the norm on $\Omega_N$ such that for any $x \in \Omega_N$, $\|x\|_\sigma(A)$ is the smallest non-negative number such that $|g(x)| \leq \|x\|_\sigma(A) \sigma(g)$ for any $g \in A$, where $\sigma(g) = \sup_{x \in \Omega_N} \{|g(x)| : \|x\|_\sigma \leq 1\}$. Let

$E_N = \left( E_{ij} : z_{ij} = 1, i = 1, ..., N, j = 1, ..., J \right)$,
with $E_{ij} = \mathbb{E}[Y_{ij}] = e^{m_{ij}}/(1 + e^{m_{ij}})$, be the vector of expected responses corresponding to the observed entries. Further define a residual alike vector $R_N \in \Omega_N$ satisfying

$$[x, R_N]_\sigma = \sum_{i=1}^{N} \sum_{j \in S_j(i)} x_{ij}(Y_{ij} - E_{ij}), \quad x \in \Omega_N.$$  

Define an evaluation measure $U_N(\cdot, \cdot)$ such that for any $y, v \in \Omega_N$, $U_N(y, v) \in \Omega_N$ is defined by the equation

$$[x, U_N(y, v)]_\sigma = \sum_{i=1}^{N} \sum_{j \in S_j(i)} x_{ij}\{\sigma^2(y_{ij}) - \sigma^2_{ij}\}v_{ij}, \quad x \in \Omega_N,$$

where $\sigma^2(y_{ij}) = e^{y_{ij}}/(1 + e^{y_{ij}})^2$. Note when $y$ is equal to $M_N^*$ or when $v$ is a zero vector, then $U_N(y, v) = 0$. Further denote that $w_{i+} = \sum_{j \in S_j(i)} w_{ij}$, $w_{+j} = \sum_{i \in S_N(j)} w_{ij}$ and $w_{++} = \sum_{i=1}^{N} \sum_{j \in S_j(i)} w_{ij}$, where $S_N(j) = \{i = 1, \ldots, N : z_{ij} = 1\}$. We then extend the results in Haberman (1977) to prove the existence and consistency of the maximum likelihood estimator $\hat{M}_N$.

We first establish the existence of $\hat{M}_N$ by applying the fixed point theorems of Kantorovich and Akilov (1964). We start with constructing a function $F_N$ on $\Omega_N$ with a fixed point $\hat{M}_N$. Consider

$$F_N(y) = y + r_N(y) \quad \text{for} \quad y \in \Omega_N, \quad \text{where} \quad r_N : \Omega_N \mapsto \Omega_N$$

is defined by the equation,

$$[x, r_N(y)]_\sigma = \sum_{i=1}^{N} \sum_{j \in S_j(i)} x_{ij}\{Y_{ij} - E(y_{ij})\}, \quad x \in \Omega_N,$$

where $E(y_{ij}) = e^{y_{ij}}/(1 + e^{y_{ij}})$. Note that $F_N$ has a fixed point $\omega \in \Omega_N$ if and only if

$$\sum_{i=1}^{N} \sum_{j \in S_j(i)} x_{ij}\{Y_{ij} - E(\omega_{ij})\} = 0, \quad x \in \Omega_N.$$  

Let $P$ be the orthogonal projection onto $\Omega_N$. Let $\hat{E} = \{E(\hat{m}_{ij}) : i = 1, \ldots, N, j = 1, \ldots, J, z_{ij} = 1\}$ and $Y_z = \{Y_{ij} : i = 1, \ldots, N, j = 1, \ldots, J, z_{ij} = 1\}$. Then following from Berk (1972), $\hat{M}_N$ is a maximum likelihood estimator of $M_N^*$ if and only if $P\hat{E} = PY_z$. Hence, $\hat{M}_N$ exists if and only if $\omega$ exists. Furthermore, since the log-likelihood $l(Y_z, \cdot)$ is strictly concave, if the maximum likelihood estimator $\hat{M}_N$ of $M_N^*$ exists, then it must be unique. Therefore, if $\hat{M}_N$ exists, $\omega = \hat{M}_N$. So, we just need to verify the conditions of the fixed point theorem to show that the fixed point $\omega$ indeed exists.
The Kantorovich & Akilov’s fixed point theorem requires construction of a sequence that converges to the fixed point. Consider the sequence \( \{t_N : k = 0, 1, \ldots\} \), with \( t_{N_0} = M_N^* \) and \( t_{N(k+1)} = F_N(t_N) \) for \( k = 0, 1, \ldots \) Note that \( t_{N_1} = M_N^* + R_N \). To check whether this sequence is well-defined and converges to \( \hat{M}_N \), we need to examine the differential \( dF_N \) of \( F_N \) at \( y \in \Omega_N \). Note that for \( y + v \in \Omega_N \),

\[
[x, F_N(y + v) - F_N(y)]_\sigma = \sum_{i=1}^{N} \sum_{j \in S_j(i)} x_{ij} \sigma_{ij}^2 [v_{ij} + (\sigma_{ij}^2)^{-1} \{E(y_{ij}) - E(y_{ij} + v_{ij})\}]
\]

\[
= -[x, U_N(y, v)]_\sigma + o(v),
\]

where \( o(v)/\|v\|_\sigma \to 0 \) as \( \|v\|_\sigma \to 0 \). It follows that \( dF_Ny(v) = -U_N(y, v) \). Denote \( \|dF_Ny\|_\sigma(A) \) to be the smallest nonnegative number such that

\[
\|dF_Ny\|_\sigma(A) \leq \|dF_Ny\|_\sigma(A)\|v\|_\sigma(A), \quad v \in \Omega_N.
\]

Let \( A_p \) be the set consisting of all the point maps \( f_{ij} \) on \( \Omega_N \), i.e. \( f_{ij}(x) = x_{ij} \) for any \( x \in \Omega_N \). By Lemma 1(c) below, there exist sequences \( f_N \) and \( d_N \) such that

\[
\|dF_Ny\|_\sigma(A_p) \leq d_N\|y - M_N^*\|_\sigma(A_p) \quad \text{whenever} \quad \|y - M_N^*\|_\sigma(A_p) \leq f_N, \quad y \in \Omega_N.
\]

**Lemma 1.** Assume Conditions 1, 2 and 3 hold. If \( A_p = \{f_{ij} : i = 1, \ldots, N, j = 1, \ldots, J, z_{ij} = 1\} \) such that \( f_{ij}(x) = x_{ij} \) for \( x \in \Omega_N \). Let \( C_N = |A_p| \), the cardinality of \( A_p \). Then there exist sequences \( f_N > 0 \) and \( d_N \geq 0 \) satisfying the followings.

(a). As \( N \to \infty , f_N^2/\log C_N \to \infty \).

(b). As \( N \to \infty , f_N^2(N_s^{-1} + J_s^{-1}) \to 0 \).

(c). If \( y, v \in \Omega_N \) and \( \|y - M_N^*\|_\sigma(A_p) \leq f_N \), then there exists \( n < \infty \) such that for all \( N > n \),

\[
\|U_N(y, v)\|_\sigma(A_p) \leq d_N\|y - M_N^*\|_\sigma(A_p)\|v\|_\sigma(A_p). \quad \text{Furthermore}, \quad d_N f_N \to 0 \quad \text{as} \quad N \to \infty.
\]

As shown in Kantorovich & Akilov (1964, pages 695-711), if \( \|R_N\|_\sigma(A_p) < \frac{1}{2} f_N \) and \( d_N\|R_N\|_\sigma(A_p) < \frac{1}{2} \), then \( \hat{M}_N \) exists. By Lemma 2 below, we have \( \operatorname{pr}(\|R_N\|_\sigma(A_p) < \frac{1}{2} f_N) \to 1 \) as \( N \to \infty \). Therefore, it follows from Lemma 1(c) that with probability tending to 1, \( d_N\|R_N\|_\sigma(A_p) < \frac{1}{2} f_N d_N \to 0 \).
Lemma 2. Let $A \subset \Omega_N^\ast$. Let $C_N$ denote the cardinality of $A$. If there exist sequences $f_N > 0$ and $d_N \geq 0$ satisfying (a). $0 < C_N < \infty$ and $f_N^2 / \log C_N \to \infty$ as $N \to \infty$, (b). If $y, v \in \Omega_N$ and $\|y - M_N^\ast\|_\sigma(A) \leq f_N$, then there exists $n < \infty$ such that for all $N > n$, $\|U_N(y,v)\|_\sigma(A) \leq d_N\|y - M_N^\ast\|_\sigma(A)\|v\|_\sigma(A)$, (c). $d_N f_N \to 0$ as $N \to \infty$. Then $\text{pr}(\|R_N\|_\sigma(A) < \frac{1}{2} f_N) \to 1$ as $N \to \infty$.

Hence, the conditions of the fixed point theorem are satisfied with probability approaching 1. It then follows that the maximum likelihood estimators $\hat{M}_N$ exists with probability tending to 1. Since Condition 3 holds, as a direct consequence of Proposition 1, the corresponding maximum likelihood estimators $\hat{\theta}_i, i = 1, \ldots, N$ and $\hat{\beta}_j, j = 1, \ldots, J$ can be uniquely determined given $\hat{M}_N$. Therefore, with probability approaching 1 that they all exist, as $N \to \infty$. The first part of the theorem then follows.

Now we seek to prove the consistency results. Taking sequences $f_N$ and $d_N$ again as satisfying the results in Lemma 1 and $A = A_p$. Then both Lemmas 2 and 3 hold. From the results of Lemmas 2 and 3, it can be implied that as $N \to \infty$, with probability tending to 1 that,

$$\|\hat{M}_N - M_N^\ast\|_\sigma(A_p) = O(f_N).$$

(6)

From Haberman (1977), $\sigma(g)$ is in fact the standard deviation of $g(\hat{M}_N)$. We further note by Lemma 4 below,

$$\max_{g \in A_p} \sigma(g) \leq \tau_2^{-1} \left( N_*^{-1} + J_*^{-1} \right)^{1/2},$$

(7)

for some $0 < \tau_2 < \infty$.

Lemma 3. Assume Conditions 1, 2 and 3 hold. Let $A \subset \Omega_N^\ast$. If there exist sequences $f_N > 0$ and $d_N \geq 0$ satisfying (a). $\text{pr}(\|R_N\|_\sigma(A) < \frac{1}{2} f_N) \to 1$ as $N \to \infty$, (b). If $y, v \in \Omega_N$ and $\|y - M_N^\ast\|_\sigma(A) \leq f_N$, then there exists $n < \infty$ such that for all $N > n$, $\|U_N(y,v)\|_\sigma(A) \leq d_N\|y - M_N^\ast\|_\sigma(A)\|v\|_\sigma(A)$, (c). $d_N f_N \to 0$ as $N \to \infty$. Then, as $N \to \infty$, with probability
approaching 1 that,
\[
\left| \frac{\| \hat{M}_N - M_N^* \|_\sigma(A)}{\| R_N \|_\sigma(A)} - 1 \right| \leq d_N^{\frac{1}{2}} \to 0 \quad \text{and} \quad \| \hat{M}_N - M_N^* - R_N \|_\sigma(A) \leq d_N \| R_N \|_\sigma^2(A).
\]

**Lemma 4.** Assume Conditions 1, 2 and 3 hold and \( \sum_{i=1}^N \theta_i = 0 \), the asymptotic variance of the maximum likelihood estimator of \( m_{ij}^* \), \( \text{var}(\hat{m}_{ij}) \), for any \( i = 1, ..., N \) and \( j = 1, ..., J \), takes the form,
\[
\text{var}(\hat{m}_{ij}) = \left( \sigma_i^2 \right)^{-1} + \left( \sigma_{ij}^2 \right)^{-1} + O(N^{-1}J^{-1}) \quad \text{as} \quad N \to \infty.
\]

Then as \( N \to \infty \), we have with probability approaching 1 that
\[
\max_{i,j,z_{ij}=1} |\hat{m}_{ij} - m_{ij}^*| = \max_{i,j,z_{ij}=1} |f_{ij}(\hat{M}_N) - f_{ij}(M_N^*)|
\]
\[
= \max_{i,j,z_{ij}=1} |f_{ij}(\hat{M}_N - M_N^*)| \leq \max_{i,j,z_{ij}=1} \sigma(f_{ij}) \| \hat{M}_N - M_N^* \|_\sigma(A_p)
\]
\[
\leq \| \hat{M}_N - M_N^* \|_\sigma(A_p) \left\{ \max_{g \in A_p} \sigma(g) \right\}
\]
\[
= O \left\{ f_N \left( N^{-1} + J^{-1} \right)^{\frac{1}{2}} \right\}
\]
\[
\to 0. \quad (8)
\]

The second last line follows from (6) and (7) and the last line follows from Lemma 1(b).

To derive explicit rates of convergence for \( \| \theta - \theta^* \|_\infty \) and \( \| \hat{\beta} - \beta^* \|_\infty \), we adopt a similar approach as in the derivation of convergence of \( \max_{i,j,z_{ij}=1} |\hat{m}_{ij} - m_{ij}^*| \). In particular, for the column parameters \( \beta_j \), we consider linear functions \( g_j \in \Omega_N^* \) such that \( g_j(x) = \beta_j \). We can construct \( g_j \) as follows. The idea is to include all the row parameters \( \theta_i \) so as to use the identifiability constraint \( \sum_{i=1}^N \theta_i = 0 \). For any \( i \in S_N(j) \), we use \( m_{ij} = \theta_i - \beta_j \) in the construction. While for each \( i \in S_{N_\theta}(j) \), where \( S_{N_\theta}(j) = \{1, 2, ..., N\} \setminus S_N(j) \), by Condition 3, there must exist \( 1 \leq i_{l1}, i_{l2}, ..., i_{lk} \leq N \) and
$1 \leq j_{i1}, j_{i2}, \ldots, j_{ik} \leq J$ such that
\[ z_{i_{j_{i1}}, j_{j_{i1}}} = z_{i_{j_{i2}}, j_{j_{i2}}} = \ldots = z_{i_{j_{ik}}, j_{j_{ik}}} = 1. \]

Therefore, we can construct $g_j$ as
\[
g_j(x) = -\frac{1}{N} \left\{ \sum_{i \in S_N(j)} m_{ij} + \sum_{i \in S_{\phi(j)}} \left( m_{i,j_{j_{i1}}} - m_{i_{j_{i1}}, j_{j_{i1}}} + m_{i_{j_{i2}}, j_{j_{i2}}} - \ldots - m_{i_{j_{ik}}, j_{j_{ik}}} + m_{i_{j_{ik}}, j_{j_{ik}}} \right) \right\}
= \beta_j.
\]

Let $A_\beta = \{g_j : j = 1, \ldots, J\}$. Now consider a sequence $f_N$ satisfying the rate requirements $f_N^2 / \log J \to \infty$ and $f_N^2 N_*^{-1/2} \to 0$ as $N \to \infty$. Then by Lemma 5 below, we can pick a sequence $d_N$ satisfying Lemma 5(a) and Lemma 5(b). Furthermore, by Lemma 6 below, we know that $\sigma^2(g_j) = (\sigma_j^2)^{-1} + O\left( (N_* J_*)^{-1} \right)$ for any $g_j \in A_\beta$. Therefore, there exist positive $0 < c_1, c_2 < \infty$ and some $n$ such that for all $N > n$,
\[ c_1^{-1} N_*^{-\frac{1}{2}} < \max_{j=1, \ldots, J} \sigma(g_j) < c_2^{-1} N_*^{-\frac{1}{2}}. \]

**Lemma 5.** Assume Conditions 1, 2 and 3 hold. If $A_\beta = \{g_j : j = 1, \ldots, J\}$ such that $g_j \in \Omega_N$ and $g_j(x) = \beta_j$ for $x \in \Omega_N$. Let $C_N = |A_\beta| = J$ be the cardinality of $A_\beta$. For any positive sequence $f_N$ such that $f_N^2 / \log J \to \infty$ and $f_N^2 N_*^{-1/2} \to 0$ as $N \to \infty$, there exists a sequence $d_N \geq 0$ satisfying the followings.

(a). If $y, v \in \Omega_N$ and $\|y - M_N^*\|_\sigma(A_\beta) \leq f_N$, then there exists $n < \infty$ such that for all $N > n$, $\|U_N(y, v)\|_\sigma(A_\beta) \leq d_N \|y - M_N^*\|_\sigma(A_\beta) \|v\|_\sigma(A_\beta)$.

(b). $d_N f_N^2 \to 0$ as $N \to \infty$.

**Lemma 6.** Assume Conditions 1, 2 and 3 hold and $\sum_{i=1}^N \theta_i = 0$. The asymptotic variance of the maximum likelihood estimator of an individual column parameter, $\text{var}(\hat{\beta}_j)$, asymptotically attains
the oracle variance \((\sigma^2_{+j})^{-1}\) in the sense that

\[
\text{var}(\hat{\beta}_j) = (\sigma^2_{+j})^{-1} + O(N^{-1}J^{-1}) \quad \text{as} \quad N \to \infty.
\]

Note that by taking sequences \(f_N\) and \(d_N\) satisfying the conditions in Lemma 5 and setting \(A = A_\beta\), it can be shown easily that the results of Lemmas 2 and 3 still hold. Hence, it can be implied that as \(N \to \infty\), with probability tending to 1,

\[
\|\hat{M}_N - M^*_N\|_\sigma(A_\beta) = O(f_N).
\]

Then as \(N \to \infty\), we have with probability approaching 1 that,

\[
\max_{j=1,\ldots,J} |\hat{\beta}_j - \beta^*_j| = \max_{j=1,\ldots,J} |g_j(\hat{M}_N) - g_j(M^*_N)|
\leq \|\hat{M}_N - M^*_N\|_\sigma(A_\beta) \max_{j=1,\ldots,J} \sigma(g_j)
\leq c_2^{-1}N_*^{-\frac{1}{2}}\|\hat{M}_N - M^*_N\|_\sigma(A_\beta)
= O\left\{(\log J)^{\frac{1}{2}}N_*^{-\frac{1}{2}}\right\} \quad \text{as} \quad N \to \infty,
\]

where the last step can be implied from the results in Lemma 7 below applied to the fact that \(\|\hat{M}_N - M^*_N\|_\sigma(A_\beta) = O(f_N)\), and the rate requirement of \(f_N\) in Lemma 5, where the minimum order of \(f_N\) is determined by \(f^2_N/\log J \to \infty\) as \(N \to \infty\). Therefore,

\[
\|\hat{\beta} - \beta^*\|_\infty = O_p\left\{(\log J)^{\frac{1}{2}}N_*^{-\frac{1}{2}}\right\}.
\]

(9)

Lemma 7. Let \(a_N\) and \(c_N\) be positive sequences. As \(N \to \infty\), suppose that \(a_N = O(b_N)\), for any sequence \(b_N\) satisfying \(b_N/c_N \to \infty\). Then \(a_N = O(c_N)\) as \(N \to \infty\).

Now for the row parameters \(\theta_i\), we adopt a similar strategy by constructing linear functions
$g_i \in \Omega_N^*$ such that $g_i(x) = \theta_i$. In specific, we can construct the linear function $g_i$ as follows.

$$g_i(x) = \frac{1}{|S_J(i)|} \sum_{j \in S_J(i)} \{m_{ij} + g_j(x)\} = \frac{1}{|S_J(i)|} \sum_{j \in S_J(i)} (\theta_i - \beta_j + \beta_j) = \theta_i,$$

where $|S_J(i)|$ denotes the cardinality of $S_J(i)$. Let $A_\theta$ consist of $g_i, i = 1, \ldots, N$, i.e. $A_\theta = \{g_i : i = 1, \ldots, N\}$. Take a positive sequence $f_N$ satisfying the rate requirements $f_N^2/\log N \to \infty$ and $f_N^2 J_*^{-1} \to 0$ as $N \to \infty$, then by Lemma 8 below, we can pick a sequence $d_N$ satisfying Lemma 8(a) and Lemma 8(b). Furthermore, by Lemma 9 below, we know that $\sigma^2(g_i) = (\sigma^2_{i+})^{-1} + O\{(N_* J_*)^{-1}\}$ for any $g_i \in A_\theta$. Hence, there exist positive $0 < \gamma_1, \gamma_2 < \infty$ and such that

$$\gamma_1^{-1} J_*^{-1/2} < \max_{i=1,\ldots,N} \sigma(g_i) < \gamma_2^{-1} J_*^{-1/2}.$$

**Lemma 8.** Assume Conditions 1, 2 and 3 hold. If $A_\theta = \{g_i : i = 1, \ldots, N\}$ such that $g_i \in \Omega_N^*$ and $g_i(x) = \theta_i$ for $x \in \Omega_N$. Let $C_N = |A_\theta| = N$ be the cardinality of $A_\theta$. Then for any positive sequence $f_N$ such that $f_N^2/\log N \to \infty$ and $J_*^{-1} f_N^2 \to 0$ as $N \to \infty$, there exists a sequence $d_N \geq 0$ satisfying the followings.

(a) If $y, v \in \Omega_N$ and $\|y - M_N^*\|_\sigma(A_\theta) \leq f_N$, then there exists $n < \infty$ such that for all $N > n$, $\|U_N(y, v)\|_\sigma(A_\theta) \leq d_N \|y - M_N^*\|_\sigma(A_\theta) \|v\|_\sigma(A_\theta)$.

(b) $d_N f_N \to 0$ as $N \to \infty$.

**Lemma 9.** Assume Conditions 1, 2 and 3 hold and $\sum_{i=1}^N \theta_i = 0$, the asymptotic variance of an individual row parameter, $\text{var}(\hat{\theta}_i)$, asymptotically attains oracle variance $(\sigma^2_{i+})^{-1}$ in the sense that

$$\text{var}(\hat{\theta}_i) = (\sigma^2_{i+})^{-1} + O(N_*^{-1} J_*^{-1}) \quad \text{as} \quad N \to \infty.$$

Note that by taking sequences $f_N$ and $d_N$ satisfying the conditions in Lemma 8 and setting $A = A_\theta$, it can be implied easily that Lemmas 2 and 3 still hold. Similarly, from $\text{pr} (\|R_N\|_\sigma(A_\theta) < \frac{1}{2} f_N) \to 1$ and the results of Lemma 3, it can be implied as $N \to \infty$, we have with probability
tending to 1 that,

$$\|\hat{M}_N - M_N^*\|_\sigma(A\theta) = O(f_N).$$

It follows, as $N \to \infty$, we have with probability approaching 1 that,

$$\max_{i=1,\ldots,N} |\hat{\theta}_i - \theta_i| = \max_{i=1,\ldots,N} |g_i(\hat{M}_N) - g_i(M_N^*)|$$

$$\leq \|\hat{M}_N - M_N^*\|_\sigma(A\theta) \max_{i=1,\ldots,N} \sigma(g_i)$$

$$< \gamma^{-1} J_*^{-\frac{1}{2}} \|\hat{M}_N - M_N^*\|_\sigma(A\theta)$$

$$= O\left\{(\log N)^{\frac{1}{2}} J_*^{-\frac{1}{2}}\right\} \text{ as } N \to \infty,$$

where the last step can be implied from the results in Lemma 7 applied to the fact that with probability tending to 1, $\|\hat{M}_N - M_N^*\|_\sigma(A\theta) = O(f_N)$, and the rate requirement of $f_N$ in Lemma 8, where the minimum order of $f_N$ is determined by $f_N^2/\log N \to \infty$. It follows that,

$$\|\hat{\theta} - \theta^*\|_\infty = O_p\left\{(\log N)^{\frac{1}{2}} J_*^{-\frac{1}{2}}\right\}. \quad (10)$$

Combining (9) and (10), we have

$$\max_{i,j} |\hat{m}_{ij} - m_{ij}^*| = O_p\left\{(\log J)^{\frac{1}{2}} N_*^{-\frac{1}{2}} + (\log N)^{\frac{1}{2}} J_*^{-\frac{1}{2}}\right\}. \quad \text{Therefore,}

\text{the second part of the theorem follows.} \quad \Box$$

Proof of Theorem 2. We first seek to show $|\sigma^2(g)/\hat{\sigma}^2(g) - 1| \to 0$ as $N \to \infty$, where $\sigma^2(g) = \sigma\{g(\hat{M})\}$. Since Conditions 1, 2 and 3 hold and $\|w_g\|_1, \|\tilde{w}_g\|_1 < C$, by Lemma 10 below,

$$|\sigma^2(g) - \hat{\sigma}^2(g)| = O(N_*^{-1} J_*^{-1}) \text{ as } N \to \infty. \quad (11)$$

Hence, it follows

$$\left|\frac{\sigma^2(g)}{\hat{\sigma}^2(g)} - 1\right| = \frac{|\sigma^2(g) - \hat{\sigma}^2(g)|}{\hat{\sigma}^2(g)} \to 0 \text{ as } N \to \infty,$$

where the last step follows from (11) and the definition of $\hat{\sigma}^2(g)$.
Lemma 10. Assume Conditions 1, 2 and 3 hold and \( \sum_{i=1}^{N} \theta_i = 0 \). Consider a linear function \( g : \Omega_N \mapsto \mathbb{R} \) with \( g(x) = \sum_{i=1}^{N} h_i \theta_i + \sum_{j=1}^{J} h'_j \beta_j \). If there exists a positive \( C < \infty \) such that \( \sum_{i=1}^{N} |h_i| < C \) and \( \sum_{j=1}^{J} |h'_j| < C \), then

\[
\sigma^2(g) = \sum_{i=1}^{N} h_i^2 (\sigma_i^2)^{-1} + \sum_{j=1}^{J} h'_j^2 (\sigma_j^2)^{-1} + O(N_*^{-1} J_*^{-1}) \quad \text{as} \quad N \to \infty.
\]

Then if we can show \( \sigma(g)^{-1} \{ g(\hat{M}) - g(M^*) \} \to N(0,1) \) in distribution, the first part of the theorem would follow directly. As a direct application of Proposition 1, we can re-write function \( g \) on \( \Omega_N \) using \( [\cdot, \cdot]_\sigma \) as follows. Let \( c_N \in \Omega_N \) be defined by the equation

\[
g(x) = [c_N, x] = \sum_{i=1}^{N} \sum_{j \in S_j(i)} c_{ij} x_{ij}\sigma_{ij}^2, \quad x \in \Omega_N.
\]

Then we can express,

\[
g(\hat{M}_N) - g(M^*_N) = [c_N, \hat{M}_N - M^*_N] = [c_N, \hat{M}_N - M^*_N - R_N] + [c_N, R_N]. \quad (12)
\]

Recall that \( \sigma(g) = \sup_{x \in \Omega_N} \{ [c_N, x]_\sigma : \|x\|_\sigma \leq 1 \} \), the supremum is attained at \( x = c_N/\|c_N\|_\sigma \), so \( \sigma(g) = \|c_N\|_\sigma \). We consider two possible cases, \( w_g = 0 \) in case 1 and \( w_g \neq 0 \) in case 2, and we seek to prove the result of the theorem hold under both cases separately.

We first consider case 1. Similar as in the proof of Theorem 1, we consider a set \( A_\beta \) consisting of linear functions \( g_j \in \Omega_N^* \) on \( \Omega_N \) such that \( g_j(x) = \beta_j \) with \( A_\beta = \{ g_j : j = 1, \ldots, J \} \). We now pick a positive sequence \( f_N \) satisfying \( f_N^2/\log J \to \infty \) and \( f_N^2 N_*^{-1/2} \to 0 \) as \( N \to \infty \). Then by Lemma 5, we can pick a sequence \( d_N \geq 0 \) satisfying Lemma 5(a) and Lemma 5(b). Furthermore, it can be implied that Lemmas 2 and 3 still hold by taking \( A = A_\beta \). Moreover, Lemma 6 implies there exist \( 0 < \gamma_1, \gamma_2 < \infty \) and some \( n \) such that for all \( N > n \),

\[
\gamma_1^{-1} N_*^{-1/2} < \sigma(g_j) < \gamma_2^{-1} N_*^{-1/2}, \quad g_j \in A_\beta. \quad (13)
\]
Now for any $x \in \Omega_N$,

$$|g(x)| = |\tilde{w}_g^T \beta|$$

$$\leq \|\tilde{w}_g\|_1 \max_{j=1,\ldots,d} \{|\beta_j|\}$$

$$\leq C \max_{g_j \in A_\beta} \{ |g_j(x)| \}$$

$$= C \max_{g_j \in A_\beta} \left\{ \frac{|g_j(x)|}{\sigma(g_j)} \sigma(g_j) \right\}$$

$$\leq C \left\{ \max_{g_j \in A_\beta} \frac{|g_j(x)|}{\sigma(g_j)} \right\} \max_{g_j \in A_\beta} \sigma(g_j)$$

$$= C \|x\|_{\sigma(A_\beta)} \max_{g_j \in A_\beta} \sigma(g_j)$$

$$\leq C \gamma_2^{-1} N_*^{-\frac{1}{2}} \|x\|_{\sigma(A_\beta)},$$  \hspace{1cm} (14)

where the second last step follows from the definition of $\| \cdot \|_{\sigma(A_\beta)}$ and the last step follows from (13). Since case 1 assumes $w_g = 0$, so $g(M) \neq 0$ implies $\tilde{w}_g \neq 0$. Then as a direct consequence of Lemma 10, there exists some $0 < \gamma_3 < \infty$ such that for all $N > n$,

$$\sigma(g) \geq \gamma_3 N_*^{-\frac{1}{2}}.$$  \hspace{1cm} (15)

As a result of (14), we have

$$\left| \left[ c_N, \hat{M}_N - M_N - R_N \right]_{\sigma} \right| \leq C \gamma_2^{-1} N_*^{-\frac{1}{2}} \| \hat{M}_N - M_N - R_N \|_{\sigma(A_\beta)}.$$  \hspace{1cm} (16)

Note that from (12),

$$\frac{g(\hat{M}_N) - g(M_N^*)}{\sigma(g)} = \frac{[c_N, \hat{M}_N - M_N^* - R_N]_{\sigma}}{\sigma(g)} + [c_N, R_N]_{\sigma}.$$
Rearrange gives as $N \to \infty$, with probability tending to 1 that,

$$
\left| \frac{g(\hat{M}_N) - g(M_N^*)}{\sigma(g)} - \left[ c_N, R_N \right]_\sigma \right| = \left| \frac{[c_N, \hat{M}_N - M_N^* - R_N]_\sigma}{\sigma(g)} \right|
$$

$$
\leq \frac{C\gamma_2^{-1}N_*^{-2}}{\sigma(g)} \| \hat{M}_N - M_N^* - R_N \|_{\sigma(A_\beta)}
$$

$$
\leq C\gamma_2^{-1}\gamma_3^{-1}d_N [\|R_N\|_{\sigma(A_\beta)}]^2
$$

$$
\leq \frac{1}{4}C\gamma_2^{-1}\gamma_3^{-1}d_N f_N^2
$$

$$
\to 0, \quad (17)
$$

where the second line follows from (16), the third line can be obtained from (15) and Lemma 3, the second last line can be implied by Lemma 2 and the last line follows from Lemma 5. Hence, it turns out that it suffices to show $[c_N, R_N]_\sigma/\sigma(g) \to N(0, 1)$. Write $Z_N = [c_N, R_N]_\sigma/\sigma(g) = \sum_{i=1}^N \sum_{j \in S_j(i)} \{c_{ij}(Y_{ij} - E_{ij})\}/\|c_N\|_\sigma$ for simplicity. The strategy is to show the moment generating function of $Z_N$, denoted as $G_{Z_N}(t)$, converges to $\exp\{t^2/2\}$, the moment generating function of the standard Gaussian. Write $c'_{ij} = c_{ij}/\|c_N\|_\sigma = c_{ij}/\sigma(g)$ for simplicity. We consider the log moment generating function of $Z_N$,

$$
\log G_{Z_N}(t) = \log \mathbb{E}[e^{tZ_N}] = \log \mathbb{E} \left[ \exp \left\{ \frac{t}{\sigma(g)} \sum_{i=1}^N \sum_{j \in S_j(i)} c_{ij}(Y_{ij} - E_{ij}) \right\} \right]
$$

$$
= -t \sum_{i=1}^N \sum_{j \in S_j(i)} c'_{ij}E_{ij} + \log \prod_{i=1}^N \prod_{j \in S_j(i)} \mathbb{E} \left\{ \exp(tc'_{ij}Y_{ij}) \right\}
$$

$$
= -t \sum_{i=1}^N \sum_{j \in S_j(i)} c'_{ij}E_{ij} + \sum_{i=1}^N \sum_{j \in S_j(i)} \log \mathbb{E} \left\{ \exp(tc'_{ij}Y_{ij}) \right\}
$$

$$
= \sum_{i=1}^N \sum_{j \in S_j(i)} \left[ \log \left\{ 1 + \exp(m_{ij}^*) \right\}^{-1} - \log \left\{ 1 + \exp(tc'_{ij} + m_{ij}^*) \right\}^{-1} - tc'_{ij}E_{ij} \right]
$$

$$
= \sum_{i=1}^N \sum_{j \in S_j(i)} \left[ \log \left\{ h(m_{ij}^*) \right\} - \log \left\{ h(tc'_{ij} + m_{ij}^*) \right\} - tc'_{ij}E_{ij} \right], \quad (18)
$$

where $h(m_{ij}) = \{1 + \exp(m_{ij})\}^{-1}$. We can then apply Taylor expansion to $\log\{h(tc'_{ij} + m_{ij}^*)\}$ about
$m^*_{ij}$. For some $t' = \alpha t$ with $0 < \alpha < 1$,

$$\log\{h(tc'_{ij} + m^*_{ij})\} = \log\{h(m^*_{ij})\} - E_{ij}tc'_{ij} - \frac{t^2}{2} c^2_{ij}\sigma^2(m^*_{ij} + t'c'_{ij}).$$

Substitute into Equation (18),

$$\log G_N(t) = \frac{t^2}{2} \sum_{i=1}^{N} \sum_{j\in S_j(i)} c^2_{ij}\sigma^2(m^*_{ij} + t'c'_{ij}), \quad \|tc'_{ij}\|_\sigma(A_\beta) \leq f_N. \tag{19}$$

With $\|c'_{N}\|_\sigma = \|c_N\|_\sigma/\|c_N\|_\sigma = 1$, the summation term in (19) can be re-expressed as follows,

$$\sum_{i=1}^{N} \sum_{j\in S_j(i)} c^2_{ij}\sigma^2(m^*_{ij} + t'c'_{ij}) = \sum_{i=1}^{N} \sum_{j\in S_j(i)} c^2_{ij}\{\sigma^2(m^*_{ij} + t'c'_{ij}) - \sigma^2_{ij} + \sigma^2_{ij}\}$$

$$= \|c'_{N}\|^2_\sigma + \sum_{i=1}^{N} \sum_{j\in S_j(i)} c^2_{ij}\{\sigma^2(m^*_{ij} + t'c'_{ij}) - \sigma^2_{ij}\}$$

$$= 1 + \sum_{i=1}^{N} \sum_{j\in S_j(i)} c^2_{ij}\{\sigma^2(m^*_{ij} + t'c'_{ij}) - \sigma^2_{ij}\}.$$

Note that

$$\sum_{i=1}^{N} \sum_{j\in S_j(i)} c^2_{ij}\{\sigma^2(m^*_{ij} + t'c'_{ij}) - \sigma^2_{ij}\} = \frac{1}{\sigma(g)} \sum_{i=1}^{N} \sum_{j\in S_j(i)} c_{ij}\{\sigma^2(m^*_{ij} + t'c'_{ij}) - \sigma^2_{ij}\} c'_{ij}$$

$$= \frac{1}{\sigma(g)} g\{U_N(M_N^* + t'c'_{N}, c'_{N})\}$$

$$\leq \frac{C\gamma_2^{-1} N_*^{-\frac{1}{2}}}{\sigma(g)} \|U_N(M_N^* + t'c'_{N}, c'_{N})\|_\sigma(A_\beta)$$

$$\leq \frac{C\gamma_2^{-1} N_*^{-\frac{1}{2}}}{\sigma(g)} d_N \|t'c'_{N}\|_\sigma(A_\beta) \|c'_{N}\|_\sigma(A_\beta)$$

$$\leq \frac{C\gamma_2^{-1} N_*^{-\frac{1}{2}}}{\sigma(g)} d_N f_N$$

$$\leq C\gamma_2^{-1} \gamma_3^{-1} d_N f_N$$

$$\to 0 \quad \text{as} \quad N \to \infty.$$
Therefore, \( \log G_{Z_N}(t) \to t^2/2 \) as \( N \to \infty \).

Now consider case 2. We adopt a similar strategy to derive asymptotic normality as in case 1. Define set \( A_{\theta, \beta} \) to consist of linear functions \( g_i, g'_j \in \Omega_N^* \) on \( \Omega_N \) such that \( g_i(x) = \theta_i \) and \( g'_j(x) = \beta_j \), with \( A_{\theta, \beta} = \{ g_i, g'_j : i = 1, \ldots, N, j = 1, \ldots, J \} \). The explicit forms of \( g_i \) and \( g'_j \) can be found in the proof of Theorem 1.

From now onwards, we take sequences \( f_N \) and \( d_N \) as satisfying the conditions in Lemma 11 below. Note it can be implied that with such \( f_N \) and \( d_N \), Lemmas 2 and 3 still hold by taking \( A = A_{\theta, \beta} \). From Lemmas 6 and 9, we know that for any \( f \in A_{\theta, \beta} \), there exist \( 0 < c_1, c_2 < \infty \) and some \( n \) such that for all \( N > n \),

\[
c_1^{-1} N_{\ast}^{-\frac{1}{2}} < \sigma(f) < c_2^{-1} J_{\ast}^{-\frac{1}{2}}.
\]

**Lemma 11.** Assume Conditions 1, 2 and 3 hold and \( J_{\ast}^{-2} N_{\ast}(\log N)^2 \to 0 \) as \( N \to \infty \). If \( A_{\theta, \beta} = \{ g_i, g'_j : i = 1, \ldots, N, j = 1, \ldots, J \} \) such that \( g_i, g'_j \in \Omega_N^* \), and \( g_i(x) = \theta_i \) and \( g'_j(x) = \beta_j \) for \( x \in \Omega_N \). Let \( C_N = |A_{\theta, \beta}| \), the cardinality of \( A_{\theta, \beta} \). Then there exist sequences \( f_N > 0 \) and \( d_N \geq 0 \) satisfying the followings.

(a). As \( N \to \infty \), \( f_N^2 / \log C_N \to \infty \).

(b). If \( y, v \in \Omega_N \) and \( \| y - M_N^* \|_{\sigma(A_{\theta, \beta})} \leq f_N \), then there exists \( n < \infty \) such that for all \( N > n \),

\[
\| U_N(y, v) \|_{\sigma(A_{\theta, \beta})} \leq d_N \| y - M_N^* \|_{\sigma(A_{\theta, \beta})} \| v \|_{\sigma(A_{\theta, \beta})}.
\]

Furthermore, \( d_N f_N^2 \to 0 \) as \( N \to \infty \).

Now for any \( x \in \Omega_N \),

\[
|g(x)| = |w_\beta^T \theta + \tilde{w}_\beta^T \beta|
\leq \left( \| w_{g} \|_{1} + \| \tilde{w}_{g} \|_{1} \right) \max_{i = 1, \ldots, N, j = 1, \ldots, J} \{ |\theta_i|, |\beta_j| \}
= \left( \| w_{g} \|_{1} + \| \tilde{w}_{g} \|_{1} \right) \max_{f \in A_{\theta, \beta}} \{ |f(x)| \}
= \left( \| w_{g} \|_{1} + \| \tilde{w}_{g} \|_{1} \right) \max_{f \in A_{\theta, \beta}} \left\{ \frac{|f(x)|}{\sigma(f)} \sigma(f) \right\}
= \left( \| w_{g} \|_{1} + \| \tilde{w}_{g} \|_{1} \right) \max_{f \in A_{\theta, \beta}} \left\{ \frac{|f(x)|}{\sigma(f)} \right\} \max_{f \in A_{\theta, \beta}} \sigma(f)
\leq 2Cc_2^{-1} J_{\ast}^{-\frac{1}{2}} \| x \|_{\sigma(A_{\theta, \beta})}.
\]

(21)
where the last step follows from the definition of $\| \cdot \|_\sigma(A_{\theta,\beta})$, (20) and the assumption that $\|w_g\|_1, \|\tilde{w}_g\|_1 < C$. Further note that since $w_g \neq 0$, as a direct consequence of Lemma 10, there exists some $0 < c_3 < \infty$ such that for all $N > n$,

$$\sigma(g) \geq c_3 J_*^{-\frac{1}{2}}. \quad (22)$$

As a result of (21),

$$\left| \left[ c_N, \hat{M}_N - M^*_N - R_N \right]_\sigma \right| \leq 2C_c^{-1} J_*^{-\frac{1}{2}} \| \hat{M}_N - M^*_N - R_N \|_\sigma(A_{\theta,\beta}).$$

Again, we have

$$\frac{g(\hat{M}_N) - g(M^*_N)}{\sigma(g)} = \left[ c_N, \hat{M}_N - M^*_N - R_N \right]_\sigma + \left[ c_N, R_N \right]_\sigma$$

As $N \to \infty$, re-arrange gives with probability tending to 1 that,

$$\left| \frac{g(\hat{M}_N) - g(M^*_N)}{\sigma(g)} - \frac{\left[ c_N, R_N \right]_\sigma}{\sigma(g)} \right| = \left| \frac{[c_N, \hat{M}_N - M^*_N - R_N]_\sigma}{\sigma(g)} \right|
\leq \frac{2C_c^{-1} J_*^{-\frac{1}{2}}}{\sigma(g)} \| \hat{M}_N - M^*_N - R_N \|_\sigma(A_{\theta,\beta})
\leq \frac{2C_c^{-1} J_*^{-\frac{1}{2}}}{\sigma(g)} d_N \left[ \| R_N \|_\sigma(A_{\theta,\beta}) \right]^2
\leq \frac{1}{2} C_c^{-1} c_3^{-1} d_N f_N^2
\to 0, \quad (23)$$

Again, we can denote $Z_N = [c_N, R_N]_\sigma / \sigma(g)$ for notation simplicity. Similar as in case 1, we just need to show $Z_N \to N(0,1)$. We consider the log moment generating function of $Z_N$, denoted as $\log G_{Z_N}(t)$. Write $c'_{ij} := c_{ij} / \sigma(g)$. Then similarly as in the proof for case 1, we obtain

$$\log G_{Z_N}(t) = \frac{t^2}{2} \sum_{i=1}^{N} \sum_{j \in S_j(i)} c'^2_{ij} \sigma^2(m_{ij}^* + t' c'_{ij}), \quad \| t' c'_{N} \|_\sigma(A_{\theta,\beta}) \leq f_N,$$
where,
\[
\sum_{i=1}^{N} \sum_{j \in S_j(i)} c_{ij}^{2} \sigma^{2}(m_{ij}^* + t'c_{ij}^*) = 1 + \sum_{i=1}^{N} \sum_{j \in S_j(i)} c_{ij}^{2} \{ \sigma^{2}(m_{ij}^* + t'c_{ij}^*) - \sigma_{ij}^{2} \}. 
\]

Note that
\[
\sum_{i=1}^{N} \sum_{j \in S_j(i)} c_{ij}^{2} \{ \sigma^{2}(m_{ij}^* + t'c_{ij}^*) - \sigma_{ij}^{2} \} = \frac{1}{\sigma(g)} \sum_{i=1}^{N} \sum_{j \in S_j(i)} c_{ij} \{ \sigma^{2}(m_{ij}^* + t'c_{ij}^*) - \sigma_{ij}^{2} \} c_{ij}^{'} = \frac{1}{\sigma(g)} g \{ U_{N}(M_{N}^* + t'c_{N}^{'}) \} 
\[
\leq \frac{2C_{2}^{-1} J_{\ast}^{-\frac{1}{2}}}{\sigma(g)} \| U_{N}(M_{N}^* + t'c_{N}^{'}) \|_{\sigma}(A_{\theta, \beta}) 
\]
\[
\leq \frac{2C_{2}^{-1} J_{\ast}^{-\frac{1}{2}}}{\sigma(g)} \| t'c_{N}^{'}) \|_{\sigma}(A_{\theta, \beta}) \| c_{N}^{'}) \|_{\sigma}(A_{\theta, \beta}) 
\]
\[
\leq \frac{2C_{2}^{-1} J_{\ast}^{-\frac{1}{2}}}{\sigma(g)} d_{N} f_{N} 
\leq 2C_{2}^{-1} c_{3}^{-1} d_{N} f_{N} 
\rightarrow 0 \text{ as } N \rightarrow \infty.
\]

The second line follows from \(U_{ij}(M_{N}^* + t'c_{N}^{'}, c_{N}^{'}) = (\sigma_{ij}^{2})^{-1} \{ \sigma(m_{ij}^* + t'c_{ij}^{'}) - \sigma_{ij}^{2} \} c_{ij}^{'}\). The third last step follows from \(\| c_{N}^{'}) \|_{\sigma}(A_{\theta, \beta}) \leq \| c_{N}^{'}) \|_{\sigma} = 1\) and the last step can be implied from Lemma 11(b). Therefore, \(\log G_{Z_{N}}(t) \rightarrow t^2 \frac{2}{\gamma} \text{ as } N \rightarrow \infty\). Hence, the first part of the theorem follows.

Now we seek to prove the second part of the theorem. The strategy is to show \(\frac{\hat{\sigma}^{2}(g) - \sigma^{2}(g)}{\hat{\sigma}^{2}(g)} \rightarrow 0 \text{ in probability as } N \rightarrow \infty\). Consider
\[
\frac{|\hat{\sigma}^{2}(g) - \sigma^{2}(g)|}{\hat{\sigma}^{2}(g)} = \left| \sum_{i=1}^{N} w_{gi}^{2} \{ (\hat{\sigma}_{i+}^{2})^{-1} - (\sigma_{i+}^{2})^{-1} \} + \sum_{j=1}^{J} \tilde{w}_{gj}^{2} \{ (\hat{\sigma}_{+j}^{2})^{-1} - (\sigma_{+j}^{2})^{-1} \} \right| 
\leq \left| \sum_{i=1}^{N} w_{gi}^{2} \frac{\sigma_{i+}^{2} - \sigma_{i+}^{*}^{2}}{(\sigma_{i+}^{*})^{2}(\sigma_{i+})^{2}} \right| + \sum_{j=1}^{J} \tilde{w}_{gj}^{2} \left| \frac{\sigma_{+j}^{2} - \sigma_{+j}^{*}^{2}}{(\sigma_{+j}^{*})^{2}(\sigma_{+j})^{2}} \right| 
\leq \left| \sum_{i=1}^{N} w_{gi}^{2} \left| \sum_{j \in S_{j}(i)} |c_{ij}^{2} - \sigma_{ij}^{2}| \right| \right| + \sum_{j=1}^{J} \tilde{w}_{gj}^{2} \left| \sum_{j \in S_{j}(i)} |c_{ij}^{2} - \sigma_{ij}^{2}| \right| \right| \right| \right| \right| \right| 
\leq \left| \sum_{i=1}^{N} w_{gi}^{2} \left| \sum_{j \in S_{j}(i)} |c_{ij}^{2} - \sigma_{ij}^{2}| \right| \right| + \sum_{j=1}^{J} \tilde{w}_{gj}^{2} \left| \sum_{j \in S_{j}(i)} |c_{ij}^{2} - \sigma_{ij}^{2}| \right| \right| \right| \right| \right| \right|. \quad (24)
\]
Since $m^*_{ij}, \hat{m}_{ij} \in \mathbb{R}$, $0 < \sigma^2_{ij}, \hat{\sigma}^2_{ij} < 1$. Note that there exist $0 < c_4, c_5 < \infty$ that

$$\sigma^2_{i+}, \hat{\sigma}^2_{i+} > c_4 J_*, \quad \sigma^2_{+j}, \hat{\sigma}^2_{+j} > c_5 N_*.$$

Further note that there exists a positive $c_6 < \infty$ such that

$$\max_{i,j,z=1} |\sigma^2_{iz} - \hat{\sigma}^2_{iz}| \leq c_6 \max_{i,j,z=1} |m^*_{ij} - \hat{m}_{ij}| = o_p(1), \quad \text{as } N \to \infty.$$ 

where the last line follows from (8). It follows

\begin{align}
\sum_{j \in S_N(i)} |\sigma^2_{ij} - \hat{\sigma}^2_{ij}| \leq c_6 \sum_{i,j,z=1} |m^*_{ij} - \hat{m}_{ij}| = o_p(1), \quad \text{as } N \to \infty.
\end{align}

Moreover, we note that $\|w^g\|_1, \|\tilde{w}^g\|_1 < C$ implies that $\sum_{i=1}^N w^2_{gi} < c_7$ and $\sum_{j=1}^J \tilde{w}^2_{gj} < c_7$ for some $c_7 < \infty$. From (24), it can be implied that

$$\frac{\sigma^2(g) - \hat{\sigma}^2(g)}{\hat{\sigma}^2(g)} = o_p(1), \quad \text{as } N \to \infty,$$

where the above result follows from (25), (26) and the assumption that $g(x) \neq 0$ for any $x \in \Omega_N$.

Since we have shown $\hat{\sigma}(g)^{-1} \{g(\hat{M}) - g(M^*)\} \to N(0, 1)$ in distribution in the first part of the proof, it follows that $\hat{\sigma}(g)^{-1} \{g(\hat{M}) - g(M^*)\} \to N(0, 1)$ in distribution as $N \to \infty$. \hfill \Box

**Proof of Proposition 1.** We prove the first part of the proposition by direct construction; in particular, we find the solutions for $\theta$ and $\beta$, respectively, given equations $\sum_{i=1}^N \theta_i = 0$ and $\theta_i - \beta_j = m_{ij}$, $i = 1, ..., N, j = 1, ..., J$, for which $z_{ij} = 1$. We first construct the solution for $\beta_j, j = 1, ..., J$. The idea is to include all the row parameters $\theta_i$ so that we can apply the constraint $\sum_{i=1}^N \theta_i = 0$. Denote $S_J(i) = \{j = 1, ..., J : z_{ij} = 1\}$, $S_N(j) = \{i = 1, ..., N : z_{ij} = 1\}$, and $S_{N_\phi}(j) = \{1, 2, ..., N \} \setminus S_N(j)$. Then for any $i \in S_N(j)$, we use $m_{ij} = \theta_i - \beta_j$ in the construction. While for each $i \in S_{N_\phi}(j)$,
applying Condition 3, there must exist $1 \leq i_1, i_2, \ldots, i_k \leq N$ and $1 \leq j_1, j_2, \ldots, j_k \leq J$ such that

$$z_{i_1,j_1} = z_{i_1,j_1} = z_{i_1,j_2} = \ldots = z_{i_k,j_k} = z_{i_k,j} = 1,$$

with

$$m_{i_1,j_1} - m_{i_1,j_1} + m_{i_1,j_2} - m_{i_2,j_2} + \ldots - m_{i_k,j_k} + m_{i_k,j}$$

$$= \left( \theta_i - \beta_{j_1} \right) - \left( \theta_{i_1} - \beta_{j_1} \right) + \left( \theta_{i_2} - \beta_{j_2} \right) - \left( \theta_{i_2} - \beta_{j_2} \right) + \ldots - \left( \theta_{i_k} - \beta_{j_k} \right) + \left( \theta_{i_k} - \beta_j \right)$$

$$= \theta_i - \beta_j.$$

Therefore, the solution for $\beta_j$ is simply

$$\beta_j = -\frac{1}{N} \left\{ \sum_{i \in S_N(j)} m_{ij} \right. \left. + \sum_{i \in S_{N_j}(j)} \left( m_{i,j_1} - m_{i_1,j_1} + m_{i_1,j_2} - m_{i_2,j_2} + \ldots - m_{i_k,j_k} + m_{i_k,j} \right) \right\}.$$

To find solution for $\theta_i$,

$$\theta_i = \frac{1}{|S_{f_j}(i)|} \sum_{j \in S_{f_j}(i)} \left[ m_{ij} - \frac{1}{N} \left\{ \sum_{i' \in S_N(j)} m_{i',j} \right. \left. + \sum_{i' \in S_{N_j}(j)} \left( m_{i',j_1} - m_{i_1',j_1} + m_{i_1',j_2} - m_{i_2',j_2} + \ldots - m_{i_k',j_k} + m_{i_k',j} \right) \right\} \right],$$

where $|S_{f_j}(i)|$ denotes the cardinality of $S_{f_j}(i)$. This concludes the proof for the first part of the proposition.

We can view the row parameters and column parameters as a bipartite graph $G$, with one part consisting of row parameters as nodes (denoted as $\{i = 1, \ldots, N\}$ for simplicity) and the other consisting of column parameters as nodes (denoted as $\{j = 1, \ldots, J\}$ for simplicity). If $z_{ij} = 1$, then there is an edge connecting $i$ and $j$ in $G$. For the second part of the proposition, note if Condition 3 is not satisfied, then there exists at least one pair of $(i, j)$ such that there does not exist a path connecting them in graph $G$. This means that claim: $G$ can be separated into at least two sub-graphs. Denote the two sub-graphs by $G_1$ and $G_2$ respectively. The above claim can be proved by
a contradiction argument as follows. Suppose not, then there exist either $i'_1 \in G_1$ and $j'_2 \in G_2$ with $z_{i'_1 j'_2} = 1$, or $j'_1 \in G_1$ and $i'_2 \in G_2$ with $z_{i'_2 j'_1} = 1$. By assumption there must exist a path connecting any two nodes within each of the two sub-graphs, otherwise we could split $G$ into two sub-graphs. Therefore, there must exist a path connecting the pair $(i,j)$. A contradiction.

Now, denote \{\theta_{i_1}, \beta_{j_1} : 1 \leq i_1 \leq N, 1 \leq j_1 \leq J\} and \{\theta_{i_2}, \beta_{j_2} : 1 \leq i_2 \leq N, 1 \leq j_2 \leq J\} as the values associated with the nodes in $G_1$ and in $G_2$ respectively and together also serving as a solution set satisfying $\sum_{i=1}^{N} \theta_{i} = 0$ and $\theta_{i} - \beta_{j} = m_{ij}, i = 1,...,N, j = 1,...,J, z_{ij} = 1$. Let $n_{i_1}$ and $n_{i_2}$ denote the number of row parameters in $G_1$ and in $G_2$ respectively. Let $\tau = n_{i_1}/n_{i_2}$. For any constant $a$, let $\tilde{\theta}_{i_1} = \theta_{i_1} + a, \tilde{\beta}_{j_1} = \beta_{j_1} + a$ and $\tilde{\theta}_{i_2} = \theta_{i_2} - \tau a, \tilde{\beta}_{j_2} = \beta_{j_2} - \tau a$. We can check easily that $(\tilde{\theta}, \tilde{\beta})$ is also a solution to the system but $(\tilde{\theta}, \tilde{\beta}) \neq (\theta, \beta)$. This concludes the proof for the second part of the proposition. \hfill $\square$

### A.2 Proofs of Supporting Lemmas

We first give some intuition on how to obtain the approximation formula for $\sigma^2(g)$, as summarized in Lemmas 12, 13 and 14 below. Lemmas 12, 13 and 14 will be used in the proofs of other supporting lemmas, which will be given later in this section.

First note that it is a property of the exponential family that $\sigma(g) = \sup_{x \in \Omega_N} \{|g(x)| : \|x\|^2_{\beta} \leq 1\}$ (see e.g. page 823 of Haberman (1977)). $\sigma^2(g)$ can be viewed as the solution to a constrained quadratic programming problem, i.e.

$$\max_{\theta, \beta} \left\{ \sum_{i=1}^{N} \sum_{j \in S_J(i)} w_{ij} (\theta_{i} - \beta_{j}) \right\}^2 \quad \text{such that} \quad \sum_{i=1}^{N} \sum_{j \in S_J(i)} \sigma_{ij}^2 (\theta_{i} - \beta_{j})^2 \leq 1, \sum_{i=1}^{N} \theta_{i} = 0. \quad (27)$$

An explicit form is often difficult to derive, so an approximation is desired for both implementation and inference purposes. We consider a three-way decomposition of the coefficients of $g$ that lies in the constrained solution space, and convert this quadratic programming to a linear system from which $\sigma^2(g)$ can be solved. The results are summarized in Lemma 12 below.

**Lemma 12.** Define a vector $d(g) = \{d_{ij}(g) : i = 1,...,N, j = 1,...,J, z_{ij} = 1, d_{ij}(g) \in \mathbb{R}\}$ with a three-way decomposition $d_{ij}(g) = b(g) + f_i(g) + m_{ij}(g)$, such that $[d(g), x]_{\sigma} = g(x)$ for $x \in \Omega_N$ and
\( f_i(g), m_j(g) \) satisfying

\[
\sum_{i=1}^{N} \sigma_{i+}^2 f_i(g) = 0, \quad (28)
\]

\[
\sum_{j=1}^{J} \sigma_{j+}^2 m_j(g) = 0. \quad (29)
\]

Then, we have

\[
\sigma^2(g) = b^2(g) \sigma_{++}^2 + \sum_{i=1}^{N} \sigma_{i+}^2 f_i^2(g) + \sum_{j=1}^{J} \sigma_{j+}^2 m_j^2(g) + 2 \sum_{i=1}^{N} \sum_{j \in S_J(i)} \sigma_{ij}^2 f_i(m_j g) y + \sum_{j=1}^{J} \sum_{i \in S_N(j)} \sigma_{ij}^2 m_j(g) y. \quad (30)
\]

**Proof.** Note \( \sigma^2(g) \) is a solution to the quadratically constrained quadratic programming problem (27). From Haberman (1977), the construction of \( d(g) \) in the lemma lies in the required solution space of (27). As a result, \( \sigma^2(g) \) can be expressed directly as \( \sigma^2(g) = \|d(g)\|_\sigma^2 \). We just need to find an explicit expression of \( \|d(g)\|_\sigma^2 \) in terms of \( b(g), f_i(g), m_j(g) \).

First consider \( x \in \Omega_N \) such that \( x_{ij} = y \) are identical for all \( i = 1, \ldots, N, j = 1, \ldots, J, z_{ij} = 1 \). Then in such cases,

\[
g(x) = [d(g), x]_\sigma
\]

\[
= \sum_{i=1}^{N} \sum_{j \in S_J(i)} \{b(g) + f_i(g) + m_j(g)\} \sigma_{ij}^2 y
\]

\[
= b(g) \sigma_{++}^2 y + \sum_{i=1}^{N} \left( \sum_{j \in S_J(i)} \sigma_{ij}^2 \right) f_i(g) y + \sum_{j=1}^{J} \left( \sum_{i \in S_N(j)} \sigma_{ij}^2 \right) m_j(g) y
\]

\[
= b(g) \sigma_{++}^2 y + \sum_{i=1}^{N} \sigma_{i+}^2 f_i(g) y + \sum_{j=1}^{J} \sigma_{j+}^2 m_j(g) y
\]

\[
= b(g) \sigma_{++}^2 y, \quad (31)
\]

where the last step follows from (28) and (29). Also by the original definition of \( g \), we have

\[
g(x) = \sum_{i=1}^{N} \sum_{j \in S_J(i)} w_{ij} y = w_{++} y. \quad (32)
\]
Since (31) and (32) hold for any $y$, we must have

$$b(g) = (\sigma^2_{++})^{-1}w_{++}. \quad (33)$$

Next consider $x \in \Omega_N$ such that $x_{ij} = y_i$, $y_i \in \mathbb{R}$, for any $i = 1, ..., N, j = 1, ..., J, z_{ij} = 1$, then

$$g(x) = [d(g), x]_\sigma = \sum_{i=1}^N \sum_{j \in S_J(i)} d_{ij}(g)\sigma^2_{ij}y_i = \sum_{i=1}^N y_i \left( \sum_{j \in S_J(i)} d_{ij}(g)\sigma^2_{ij} \right). \quad (34)$$

From the original definition of $g$,

$$g(x) = \sum_{i=1}^N \sum_{j \in S_J(i)} w_{ij}y_i = \sum_{i=1}^N y_i \left( \sum_{j \in S_J(i)} w_{ij} \right). \quad (35)$$

Since (34) = (35) for any $y_i$, it follows that

$$\sum_{j \in S_J(i)} d_{ij}(g)\sigma^2_{ij} = \sum_{j \in S_J(i)} w_{ij} = w_{i+}, \quad i = 1, ..., N. \quad (36)$$

Consider

$$f_i(g) + m_j(g) = d_{ij}(g) - b(g)$$

$$\sum_{j \in S_J(i)} \{ f_i(g) + m_j(g) \} \sigma^2_{ij} = \sum_{j \in S_J(i)} \{ d_{ij}(g) - b(g) \} \sigma^2_{ij}$$

$$\sigma^2_{i+}f_i(g) + \sum_{j \in S_J(i)} \sigma^2_{ij}m_j(g) = \sum_{j \in S_J(i)} d_{ij}(g)\sigma^2_{ij} - \sigma^2_{i+}b(g)$$

$$\sigma^2_{i+}f_i(g) + \sum_{j \in S_J(i)} \sigma^2_{ij}m_j(g) = w_{i+} - (\sigma^2_{++})^{-1}w_{++}\sigma^2_{i+}, \quad i = 1, ..., N, \quad (37)$$

where the last line follows from (33) and (36). Similarly, we consider $x \in \Omega_N$ such that $x_{ij} = y_j$, $y_j \in \mathbb{R}$ for any $i = 1, ..., N, j = 1, ..., J, z_{ij} = 1$, then

$$g(x) = [d(g), x]_\sigma = \sum_{j=1}^J \sum_{i \in S_N(j)} d_{ij}(g)\sigma^2_{ij}y_j$$

$$= \sum_{j=1}^J y_j \left( \sum_{i \in S_N(j)} d_{ij}(g)\sigma^2_{ij} \right). \quad (38)$$
Again by the original definition of $g$,

$$g(x) = \sum_{j=1}^{J} \sum_{i \in S_N(j)} w_{ij} y_j = \sum_{j=1}^{J} y_j \left( \sum_{i \in S_N(j)} w_{ij} \right). \quad (39)$$

Since (38) = (39) for any $y_j \in \mathbb{R}$, it follows

$$\sum_{i \in S_N(j)} d_{ij}(g) \sigma_{ij}^2 = \sum_{i \in S_N(j)} w_{ij} = w_{+j}. \quad (40)$$

Similarly,

$$f_i(g) + m_j(g) = d_{ij}(g) - b(g)$$

$$\sum_{i \in S_N(j)} \{ f_i(g) + m_j(g) \} \sigma_{ij}^2 = \sum_{i \in S_N(j)} \{ d_{ij}(g) - b(g) \} \sigma_{ij}^2$$

$$\sigma_{i,j}^2 m_j(g) + \sum_{i \in S_N(j)} \sigma_{ij}^2 f_i(g) = \sum_{i \in S_N(j)} d_{ij}(g) \sigma_{ij}^2 - \sigma_{i,j}^2 b(g)$$

$$\sigma_{i,j}^2 m_j(g) + \sum_{i \in S_N(j)} \sigma_{ij}^2 f_i(g) = w_{+j} - (\sigma_{++}^2)^{-1} w_{+} \sigma_{+j}^2, \quad j = 1, ..., J, \quad (41)$$

where the last line follows from (33) and (40). Note that all $b(g), f_i(g), m_j(g)$ can be obtained by solving a system of $N + J + 1$ linear equations from (33), (37) and (41). Now we seek to derive a simplified expression for $\|d(g)\|^2_\sigma$ in terms of $b(g), f_i(g), m_j(g)$.

$$\sigma^2(g) = \|d(g)\|^2_\sigma$$

$$= \sum_{i=1}^{N} \sum_{j \in S_j(i)} \sigma_{ij}^2 \{ b(g) + f_i(g) + m_j(g) \}^2$$

$$= b(g) \sum_{i=1}^{N} \sum_{j \in S_j(i)} \sigma_{ij}^2 \{ b(g) + f_i(g) + m_j(g) \} \quad (42)$$

$$+ \sum_{i=1}^{N} \sum_{j \in S_j(i)} f_i(g) \sigma_{ij}^2 \{ b(g) + f_i(g) + m_j(g) \} \quad (43)$$

$$+ \sum_{i=1}^{N} \sum_{j \in S_j(i)} m_j(g) \sigma_{ij}^2 \{ b(g) + f_i(g) + m_j(g) \}. \quad (44)$$
Let us consider each of these three terms separately,

\[(42) = b^2(g)\sigma^2_{i++} + b(g) \sum_{i=1}^{N} f_i(g) \left( \sum_{j \in S_N(i)} \sigma_{ij}^2 \right) + b(g) \sum_{j=1}^{J} m_j(g) \left( \sum_{i \in S_N(j)} \sigma_{ij}^2 \right)\]

\[= b^2(g)\sigma^2_{i++} + b(g) \sum_{i=1}^{N} \sigma_{i+i}^2 f_i(g) + b(g) \sum_{j=1}^{J} \sigma_{i+j}^2 m_j(g)\]

\[= b^2(g)\sigma^2_{i++}.

\[(43) = b(g) \sum_{i=1}^{N} f_i(g) \left( \sum_{j \in S_N(i)} \sigma_{ij}^2 \right) + b(g) \sum_{i=1}^{N} f_i^2(g) \left( \sum_{j \in S_N(i)} \sigma_{ij}^2 \right) + b(g) \sum_{i=1}^{N} \sum_{j \in S_N(i)} \sigma_{ij}^2 f_i(g) m_j(g)\]

\[= b(g) \sum_{i=1}^{N} f_i^2(g) + b(g) \sum_{i=1}^{N} \sum_{j \in S_N(i)} \sigma_{ij}^2 f_i(g) m_j(g)\]

\[(44) = b(g) \sum_{j=1}^{J} m_j(g) \left( \sum_{i \in S_N(j)} \sigma_{ij}^2 \right) + b(g) \sum_{j=1}^{J} m_j^2(g) \left( \sum_{i \in S_N(j)} \sigma_{ij}^2 \right) + b(g) \sum_{i=1}^{N} \sum_{j \in S_N(i)} \sigma_{ij}^2 f_i(g) m_j(g)\]

\[= b(g) \sum_{j=1}^{J} \sigma_{i+j}^2 m_j(g) + b(g) \sum_{j=1}^{J} \sigma_{i+j}^2 m_j^2(g) + b(g) \sum_{i=1}^{N} \sum_{j \in S_N(i)} \sigma_{ij}^2 f_i(g) m_j(g)\]

\[= \sum_{j=1}^{J} \sigma_{i+j}^2 m_j^2(g) + \sum_{i=1}^{N} \sum_{j \in S_N(i)} \sigma_{ij}^2 f_i(g) m_j(g)\]

Combining three terms together, the result of the lemma follows with

\[\sigma^2(g) = ||d(g)||^2_\sigma = b^2(g)\sigma^2_{i++} + \sum_{i=1}^{N} \sigma_{i+i}^2 f_i^2(g) + \sum_{j=1}^{J} \sigma_{i+j}^2 m_j^2(g) + \sum_{i=1}^{N} \sum_{j \in S_N(i)} \sigma_{ij}^2 f_i(g) m_j(g)\]

As in the proof of Lemma 12, we can solve a system of \(N + J + 1\) linear equations from (33), (37) and (41) for \(f_i(g), i = 1, \ldots, N, m_j(g), j = 1, \ldots, J\) and \(b(g)\). Then an exact expression for \(\sigma^2(g)\) can be obtained by substituting these values into (30). However, when \(N\) and \(J\) are large, it is
difficult to solve this large system of linear equations. Furthermore, to study the order of \( \sigma^2(g) \), we need an analytical form for analysis. The following set-ups are used to find an approximation for \( \sigma^2(g) \). Define \( \gamma_N > 0 \) to be the largest number such that for all \( i = 1, \ldots, N, j = 1, \ldots, J, z_{ij} = 1, \)

\[
x^2 \sigma_{ij}^2 \geq \gamma_N \left( \frac{1}{|S_J(i)|} x^2 \sigma_{i+}^2 + \frac{1}{|S_N(j)|} x^2 \sigma_{++}^2 \right), \quad x \in \mathbb{R}, \tag{45}
\]

where \( |S_J(i)| \) and \( |S_N(j)| \) are the cardinalities of \( S_J(i) \) and \( S_N(j) \) respectively. Note that there exist some \( \gamma > 0 \) such that \( \gamma_N > \gamma \) for all \( N \). For \( i = 1, \ldots, N \) and \( j = 1, \ldots, J \), further define

\[
f'_i(g) = (\sigma_{i+}^2)^{-1} w_{i+} - (\sigma_{++}^2)^{-1} w_{++}, \tag{46}
\]

\[
m'_j(g) = (\sigma_{++}^2)^{-1} w_{++} - (\sigma_{++}^2)^{-1} w_{++}, \tag{47}
\]

\[
f''_i(g) = f_i(g) - f'_i(g), \tag{48}
\]

\[
m''_j(g) = m_j(g) - m'_j(g), \tag{49}
\]

Then for \( i = 1, \ldots, N, j = 1, \ldots, J \) with \( z_{ij} = 1, \) define

\[
\hat{\sigma}_{ij}^2 = \sigma_{ij}^2 - \gamma_N \left( \frac{1}{|S_J(i)|} \sigma_{i+}^2 + \frac{1}{|S_N(j)|} \sigma_{++}^2 \right), \tag{50}
\]

\[
d'_{ij}(g) = b(g) + f'_i(g) + m'_j(g), \tag{51}
\]

\[
d''_{ij}(g) = f''_i(g) + m''_j(g) = d_{ij}(g) - d'_{ij}(g). \tag{52}
\]

By triangle inequality, (52) then implies

\[
\|d'(g)\|_\sigma - \|d''(g)\|_\sigma \leq \|d(g)\|_\sigma \leq \|d'(g)\|_\sigma + \|d''(g)\|_\sigma.
\]

We seek to use \( \|d'(g)\|_\sigma \) as an approximation to \( \sigma(g) = \|d(g)\|_\sigma \) while showing \( \|d''(g)\|_\sigma \) is a negligible term asymptotically under certain conditions. The analytical expression for \( \|d'(g)\|_\sigma \) is given in Lemma 13 below.
Lemma 13. If $d'(g)$ is defined as in (51), then

$$
\|d'(g)\|_2^2 = \sum_{i=1}^{N} w_{i+}^2 (\sigma_{i+}^2)^{-1} + \sum_{j=1}^{J} w_{++j}^2 (\sigma_{++j}^2)^{-1}
+ 2 \sum_{i=1}^{N} \sum_{j \in S_J(i)} w_{i+j}^2 \sigma_{ij}^2 (\sigma_{i+}^2)^{-1}(\sigma_{++j}^2)^{-1} - 3w_{++}^2 (\sigma_{++}^2)^{-1}.
$$

Proof. Following from the definition of $d'(g)$, we can write

$$
\|d'(g)\|_2^2 = \sum_{i=1}^{N} \sum_{j \in S_J(i)} \sigma_{ij}^2 \{b(g) + (\sigma_{i+}^2)^{-1}w_{i+} + (\sigma_{++j}^2)^{-1}w_{++j} - 2(\sigma_{++}^2)^{-1}w_{++}\}^2
$$

$$
= b(g) \sum_{i=1}^{N} \sum_{j \in S_J(i)} \sigma_{ij}^2 \{b(g) + (\sigma_{i+}^2)^{-1}w_{i+} + (\sigma_{++j}^2)^{-1}w_{++j} - 2(\sigma_{++}^2)^{-1}w_{++}\} (53)
+ \sum_{i=1}^{N} \sum_{j \in S_J(i)} \sigma_{ij}^2 (\sigma_{i+}^2)^{-1}w_{i+} \{b(g) + (\sigma_{i+}^2)^{-1}w_{i+} + (\sigma_{++j}^2)^{-1}w_{++j} - 2(\sigma_{++}^2)^{-1}w_{++}\} (54)
+ \sum_{i=1}^{N} \sum_{j \in S_J(i)} \sigma_{ij}^2 (\sigma_{++j}^2)^{-1}w_{++j} \{b(g) + (\sigma_{i+}^2)^{-1}w_{i+} + (\sigma_{++j}^2)^{-1}w_{++j} - 2(\sigma_{++}^2)^{-1}w_{++}\} (55)
- 2 \sum_{i=1}^{N} \sum_{j \in S_J(i)} \sigma_{ij}^2 (\sigma_{++j}^2)^{-1}w_{++j} \{b(g) + (\sigma_{i+}^2)^{-1}w_{i+} + (\sigma_{++j}^2)^{-1}w_{++j} - 2(\sigma_{++}^2)^{-1}w_{++}\}. (56)
$$

We evaluate each of these four terms separately. For the first term,

$$
(53) = b^2(g)\sigma_{++}^2 + b(g) \sum_{i=1}^{N} \left( \sum_{j \in S_J(i)} \sigma_{ij}^2 \right) (\sigma_{i+}^2)^{-1}w_{i+} - b(g)w_{++} + b(g) \sum_{j=1}^{J} w_{++j} - b(g)w_{++}
$$

$$
= b^2(g)\sigma_{++}^2
= (\sigma_{++}^2)^{-1}w_{++}^2,
$$
Where the last line follows from (33). Now consider the second term,

\[(54) = b(g)w_++ + \sum_{i=1}^{N} w_{i+}^2(\sigma_{i+}^2)^{-1} + \sum_{i=1}^{N} \sum_{j \in S_j(i)} \sigma_{ij}^2(\sigma_{i+}^2)^{-1}w_{i+}w_{j+} - 2(\sigma_{++}^2)^{-1}w_{++}^2\]

\[= - (\sigma_{++}^2)^{-1}w_{++}^2 + \sum_{i=1}^{N} w_{i+}^2(\sigma_{i+}^2)^{-1} + \sum_{i=1}^{N} \sum_{j \in S_j(i)} w_{i+}w_{j+}\sigma_{ij}^2(\sigma_{i+}^2)^{-1}(\sigma_{++}^2)^{-1}.\]

Now consider the third term,

\[(55) = b(g)w_++ + \sum_{i=1}^{N} \sum_{j \in S_j(i)} \sigma_{ij}^2(\sigma_{++}^2)^{-1}w_{i+}w_{j+} + \sum_{j=1}^{J} w_{j+}^2(\sigma_{++}^2)^{-1} - 2(\sigma_{++}^2)^{-1}w_{++}^2\]

\[= - (\sigma_{++}^2)^{-1}w_{++}^2 + \sum_{j=1}^{J} w_{j+}^2(\sigma_{++}^2)^{-1} + \sum_{i=1}^{N} \sum_{j \in S_j(i)} w_{i+}w_{j+}\sigma_{ij}^2(\sigma_{i+}^2)^{-1}(\sigma_{++}^2)^{-1}.\]

Now consider the last term,

\[(56) = -2b(g)w_++ - 2(\sigma_{++}^2)^{-1}w_{++}^2 - 2(\sigma_{++}^2)^{-1}w_{++}^2 + 4b(g)w_++\]

\[= -2(\sigma_{++}^2)^{-1}w_{++}^2 - 2(\sigma_{++}^2)^{-1}w_{++}^2 - 2(\sigma_{++}^2)^{-1}w_{++}^2 - 4(\sigma_{++}^2)^{-1}w_{++}^2\]

\[= -2w_{++}^2(\sigma_{++}^2)^{-1}.\]

Combining all these four terms together, we obtain

\[\|d'(g)\|_{\sigma}^2 = \sum_{i=1}^{N} w_{i+}^2(\sigma_{i+}^2)^{-1} + \sum_{j=1}^{J} w_{j+}^2(\sigma_{++}^2)^{-1}\]

\[+ 2\sum_{i=1}^{N} \sum_{j \in S_j(i)} w_{i+}w_{j+}\sigma_{ij}^2(\sigma_{i+}^2)^{-1}(\sigma_{++}^2)^{-1} - 3w_{++}^2(\sigma_{++}^2)^{-1}.\]

Hence the result of the lemma follows. \(\square\)

Lemma 14 below gives an analytical upper bound for \(\|d'(g)\|_{\sigma}\) so that we can show it is a
negligible term under certain conditions. Define

\[ l_i = - \sum_{j \in S_J(i)} w_{ij} \sigma_{ij}^2 (\sigma_{ij}^2)^{-1} + w_{ij} \sigma_{ij}^2 (\sigma_{ij}^2)^{-1}, \quad i = 1, \ldots, N, \] (57)

\[ v_j = - \sum_{i \in S_N(j)} w_{ij} \sigma_{ij}^2 (\sigma_{ij}^2)^{-1} + w_{ij} \sigma_{ij}^2 (\sigma_{ij}^2)^{-1}, \quad j = 1, \ldots, J. \] (58)

**Lemma 14.** If \( l_i \) and \( v_j \) are defined as in (57) and (58), respectively, then

\[ \|d''(g)\|_\sigma \leq \gamma_N^{-1} \left[ \sum_{i=1}^N l_i^2 (\sigma_{ii}^2)^{-1} + \sum_{j=1}^J v_j^2 (\sigma_{jj}^2)^{-1} \right]. \]

*Proof.* From the definitions of \( f''_i, m''_j, l_i \) and \( v_j \) as in (48), (49), (57) and (58), respectively, it can be easily verified that

\[ \sigma_{ii}^2 f''_i + \sum_{j \in S_J(i)} \sigma_{ij}^2 m''_j = l_i, \quad i = 1, \ldots, N, \]

\[ \sigma_{jj}^2 m''_j + \sum_{i \in S_N(j)} \sigma_{ij}^2 f''_i = v_j, \quad j = 1, \ldots, J. \]

It can be shown \( \|d''(g)\|_\sigma^2 = \sum_{i=1}^N f''_i l_i + \sum_{j=1}^J m''_j v_j \), which can be seen as follows,

\[ \sum_{i=1}^N f''_i l_i + \sum_{j=1}^J m''_j v_j = \sum_{i=1}^N f''_i \left( \sigma_{ii}^2 f''_i + \sum_{j \in S_J(i)} \sigma_{ij}^2 m''_j \right) + \sum_{j=1}^J m''_j \left( \sigma_{jj}^2 m''_j + \sum_{i \in S_N(j)} \sigma_{ij}^2 f''_i \right) \]

\[ = \sum_{i=1}^N \sigma_{ii}^2 f''_i^2 + \sum_{j=1}^J \sigma_{jj}^2 m''_j^2 + 2 \sum_{i=1}^N \sum_{j \in S_J(i)} f''_i m''_j \sigma_{ij}^2 \]

\[ = \sum_{i=1}^N \sum_{j \in S_J(i)} (f''_i + m''_j)^2 \sigma_{ij}^2 \]

\[ = \|d''(g)\|_\sigma^2. \]

Furthermore, by Rao (1973), \( \sum_{i=1}^N f''_i l_i + \sum_{j=1}^J m''_j v_j \) is the largest value of \( \left( \sum_{i=1}^N x_i l_i + \sum_{j=1}^J y_j v_j \right)^2 \),
for $x_i \in \mathbb{R}$, $i = 1, \ldots, N$ and $y_j \in \mathbb{R}$, $j = 1, \ldots, J$ such that

$$
\sum_{i \in S_N(j)} 1_{|S_N(j)|} \sigma^2_{i} x_i = 0, \quad j = 1, \ldots, J,
$$

$$
\sum_{j \in S_N(i)} 1_{|S_N(i)|} \sigma^2_{j} y_j = 0, \quad i = 1, \ldots, N,
$$

$$
D(x, y) = \sum_{i=1}^{N} \sum_{j \in S_N(i)} \sigma^2_{ij}(x_i + y_j)^2 \leq 1.
$$

Note

$$
\sum_{i=1}^{N} \sum_{j \in S_N(i)} (x_i + y_j)^2 \sigma^2_{ij}
$$

$$
= \sum_{i=1}^{N} \sum_{j \in S_N(i)} (x_i + y_j)^2 \left\{ \sigma^2_{ij} - \gamma N \left( \frac{1}{|S_N(j)|} \sigma^2_{i} + \frac{1}{|S_N(j)|} \sigma^2_{j} \right) \right\}
$$

$$
= D(x, y) - \gamma N \sum_{i=1}^{N} \sum_{j \in S_N(i)} (x_i + y_j)^2 \left\{ \frac{1}{|S_N(j)|} \sigma^2_{i} + \frac{1}{|S_N(j)|} \sigma^2_{j} \right\}
$$

$$
= D(x, y) - \gamma N \left\{ \sum_{i=1}^{N} (x_i \sigma^2_{i} + \sum_{j \in S_N(i)} \frac{1}{|S_N(j)|} x_i \sigma^2_{j}) + \sum_{j=1}^{J} (y_j \sigma^2_{j} + \sum_{i \in S_N(j)} \frac{1}{|S_N(i)|} y_j \sigma^2_{i}) \right\}
$$

$$
- 2\gamma N \sum_{j=1}^{J} y_j \left\{ \sum_{i \in S_N(j)} \frac{1}{|S_N(i)|} \sigma^2_{i} x_i \right\} - 2\gamma N \sum_{i=1}^{N} x_i \left\{ \sum_{j \in S_N(i)} \frac{1}{|S_N(j)|} \sigma^2_{j} y_j \right\}
$$

$$
= D(x, y) - \gamma N \left\{ \sum_{i=1}^{N} (x_i \sigma^2_{i} + \sum_{j \in S_N(i)} \frac{1}{|S_N(j)|} x_i \sigma^2_{j}) + \sum_{j=1}^{J} (y_j \sigma^2_{j} + \sum_{i \in S_N(j)} \frac{1}{|S_N(i)|} y_j \sigma^2_{i}) \right\}.
$$

Re-arrange gives,

$$
D(x, y) = \gamma N \left\{ \sum_{i=1}^{N} (x_i ^2 \sigma^2_{i} + \sum_{j \in S_N(i)} \frac{1}{|S_N(j)|} x_i ^2 \sigma^2_{j}) + \sum_{j=1}^{J} (y_j ^2 \sigma^2_{j} + \sum_{i \in S_N(j)} \frac{1}{|S_N(i)|} y_j ^2 \sigma^2_{i}) \right\}
$$

$$
+ \sum_{i=1}^{N} \sum_{j \in S_N(i)} (x_i + y_j)^2 \sigma^2_{ij}
$$

$$
\geq \gamma N \left\{ \sum_{i=1}^{N} x_i ^2 \sigma^2_{i} + \sum_{j=1}^{J} y_j ^2 \sigma^2_{j} \right\}.
$$
It follows that
\[
\|d''(g)\|_\sigma \leq \gamma_N^{-1}\left[\sum_{i=1}^{N} l_i^2(\sigma_i^2)^{-1} + \sum_{j=1}^{J} v_j^2(\sigma_j^2)^{-1}\right].
\]

\[\square\]

Next, we give proofs for the supporting lemmas used in the proofs of Proposition 1 and the proofs of Theorems 1 and 2.

**Lemma 1.** Assume Conditions 1, 2 and 3 hold. If \(A_p = \{f_{ij} : i = 1, ..., N, j = 1, ..., J, z_{ij} = 1\}\) such that \(f_{ij}(x) = x_{ij}\) for \(x \in \Omega_N\). Let \(C_N = |A_p|\), the cardinality of \(A_p\). There exist sequences \(f_N > 0\) and \(d_N \geq 0\) satisfying the followings.

(a). As \(N \to \infty\), \(f^2_N/\log C_N \to \infty\).

(b). As \(N \to \infty\), \(f^2_N(N_*^{-1} + J_*^{-1}) \to 0\).

(c). If \(y, v \in \Omega_N\) and \(\|y - M_N^*\|_\sigma(A_p) \leq f_N\), then there exists \(n < \infty\) such that for all \(N > n\),
\[
\|U_N(y, v)\|_\sigma(A_p) \leq d_N\|y - M_N^*\|_\sigma(A_p)\|v\|_\sigma(A_p).
\]
Furthermore, \(d_N f_N \to 0\) as \(N \to \infty\).

**Proof.** Condition 1(b) assumes \(J_*^{-1} \log N \to 0\). If \(x_N \ll y_N\) means that \(y_N^{-1}x_N \to 0\) as \(N \to \infty\), then Condition 1(b) implies that \(\log N_* \ll J_*\). Then there must exist a sequence \(f_N > 0\) such that \(\log N_* \ll f^2_N \ll J_*\).

\[
\frac{f^2_N}{\log C_N} \geq \frac{f^2_N}{\log(J_*N_*)} = \frac{f^2_N}{\log J_* + \log N_*} \geq \frac{f^2_N}{2\log N_*} \to \infty \quad \text{as} \quad N \to \infty.
\]

The first inequality follows from the fact that \(J_*N_* \leq C_N \leq J_*N_*\). The last line follows from \(\log N_* \ll f^2_N\). Therefore, the result of part (a) is satisfied. We further note

\[
f^2_N(N_*^{-1} + J_*^{-1}) \leq \frac{2f^2_N}{J_*} \to 0 \quad \text{as} \quad N \to \infty.
\]

(59)
The last line follows from $f_\hat{N}_i^2 \ll J_\ast$. Therefore, part (b) of the lemma follows. To verify part (c), first note by Lemma 4, for any point maps $f_{ij} \in A_p$, there exist $0 < \tau_1, \tau_2 < \infty$ such that for all $N > n$,

$$\tau_1^{-1} \left( N_{\ast}^{-1} + J_{\ast}^{-1} \right)^{\frac{1}{2}} < \sigma(f_{ij}) < \tau_2^{-1} \left( N_{\ast}^{-1} + J_{\ast}^{-1} \right)^{\frac{1}{2}}. \tag{60}$$

By the definition of $\| \cdot \|_{\sigma(A_p)}$, we have for any $y \in \Omega_N, f_{ij} \in A_p$,

$$|f_{ij}(y)| \leq \|y\|_{\sigma(A_p)}\sigma(f_{ij}). \tag{61}$$

It follows from (60) and (61) that for any $i = 1, \ldots, N, j = 1, \ldots, J, z_{ij} = 1$,

$$|y_{ij}| \leq \tau_2^{-1}\|y\|_{\sigma(A_p)} \left( N_{\ast}^{-1} + J_{\ast}^{-1} \right)^{\frac{1}{2}}. \tag{62}$$

Since $|\sigma^2(y_{ij}) - \sigma_{ij}^2| \leq 1$, note that there exists a positive $\tau_3 < \infty$ such that for any $y \in \Omega_N$, one has for any $i = 1, \ldots, N, j = 1, \ldots, J, z_{ij} = 1$,

$$|\sigma^2(y_{ij}) - \sigma_{ij}^2| \leq \tau_3|y_{ij} - m_{ij}^\ast|. \tag{63}$$

Since $A_p$ consists of point maps only, by the definition of $\| \cdot \|_{\sigma(A_p)}$, we have $\|U_N(y, v)\|_{\sigma(A_p)}$ is the maximum value of $|f_{ij}\{U_N(y, v)\}| / \sigma(f_{ij})$ over $f_{ij} \in A_p$. Therefore, upper bounding $\|U_N(y, v)\|_{\sigma(A_p)}$ is equivalent to upper bounding all $|U_{ij}(y, v)| / \sigma(f_{ij})$. Note that for any $i = 1, \ldots, N, j = 1, \ldots, J, z_{ij} = 1$,

$$|U_{ij}(y, v)| = \left| \sum_{i'=1}^{N} \sum_{j' \in S_J(i')} \left[ d_{i'j'}(f_{ij}) \{ \sigma^2(y_{i'j'}) - \sigma_{i'j'}^2 \} v_{i'j'} \right]\right|$$

$$\leq \sum_{i'=1}^{N} \sum_{j' \in S_J(i')} \left\{ |d_{i'j'}(f_{ij})| \{ |\sigma^2(y_{i'j'}) - \sigma_{i'j'}^2| \} \{ |v_{i'j'}| \} \right\}$$

$$\leq \sum_{i'=1}^{N} \sum_{j' \in S_J(i')} \left\{ |d_{i'j'}(f_{ij})| \{ |y_{i'j'} - m_{i'j'}^\ast| \} \{ |v_{i'j'}| \} \right\}$$

$$\leq \tau_2^{-2} \tau_3 \left( N_{\ast}^{-1} + J_{\ast}^{-1} \right) \left\{ \|y - M_N^\ast\|_{\sigma(A_p)} \|v\|_{\sigma(A_p)} \right\} \left\{ \sum_{i'=1}^{N} \sum_{j' \in S_J(i')} |d_{i'j'}(f_{ij})| \right\},$$

53
where the second last line follows from (63) and the last line follows from (62). Further note that

\[
\sum_{i' = 1}^{N} \sum_{j' \in S_j(i')} |d'_{i'j'}(f_{ij})| \leq \sum_{i' = 1}^{N} \sum_{j' \in S_j(i')} |d'_{i'j'}(f_{ij})| + \sum_{i' = 1}^{N} \sum_{j' \in S_j(i')} |d'_{i'j'}(f_{ij})|.
\]

By definition, \(d'_{i'j'}(g) = (\sigma_{i'i}^2)^{-1} w_{i'i} + (\sigma_{j'j}^2)^{-1} w_{j'j} - (\sigma_{++})^{-1} w_{++} \) for any \(g \in \Omega_N^\ast\). When \(g = f_{ij}\), \(w_{i'i} = 0 \) if \(i' \neq i\), and \(w_{i'i} = 1 \) if \(i' = i\), \(w_{j'j} = 0 \) if \(j' \neq j\), and \(w_{j'j} = 1 \) if \(j' = j\), and \(w_{++} = 1\).

Therefore, we can rewrite

\[
\sum_{i' = 1}^{N} \sum_{j' \in S_j(i')} |d'_{i'j'}(f_{ij})| = \sum_{i' = 1}^{N} \sum_{j' \in S_j(i')} |(\sigma_{i'i}^2)^{-1} w_{i'i} + (\sigma_{j'j}^2)^{-1} w_{j'j} - (\sigma_{++})^{-1} w_{++}|
\]

\[
\leq \sum_{i' = 1}^{N} \sum_{j' \in S_j(i')} (\sigma_{i'i}^2)^{-1} |w_{i'i}| + \sum_{i' = 1}^{N} \sum_{j' \in S_j(i')} (\sigma_{j'j}^2)^{-1} |w_{j'j}|
\]

\[
+ \sum_{i' = 1}^{N} \sum_{j' \in S_j(i')} (\sigma_{++})^{-1} |w_{++}|
\]

\[
= \sum_{j' \in S_j(i)} (\sigma_{i'i}^2)^{-1} + \sum_{i' \in S_{N}(j)} (\sigma_{j'j}^2)^{-1} + \sum_{j' \in S_j(i')} (\sigma_{++})^{-1} \leq \tau_4,
\]

where \(\tau_4\) is some positive constant such that \(\tau_4 < \infty\). Note also that there exists \(\tau_5 < \infty\) such that

\[
\sum_{i' = 1}^{N} \sum_{j' \in S_j(i')} |d''_{i'j'}(f_{ij})| \leq (\mathcal{J}^4 N^4)^{\frac{1}{2}} \|d''(f_{ij})\|_\sigma \leq \tau_5.
\]

As a result,

\[
\|U_N(y,v)\|_\sigma(A_p) \leq \tau_1 \tau_2 \tau_3 (\tau_4 + \tau_5) \left( N^{-1} + J^{-1} \right)^{\frac{1}{2}} \left\{ \|y - M_N\|_\sigma(A_p) \|v\|_\sigma(A_p) \right\}.
\]

Therefore, we can set \(d_N = \tau_1 \tau_2 \tau_3 (\tau_4 + \tau_5) \left( N^{-1} + J^{-1} \right)^{\frac{1}{2}}. \)

By (59), we have \(f_N(N_{\ast}^{-1} + J_{\ast}^{-1})^{1/2} \to 0 \) as \(N \to \infty\). Therefore, it follows

\[
d_N f_N = \tau_1 \tau_2 \tau_3 (\tau_4 + \tau_5) \left( N_{\ast}^{-1} + J_{\ast}^{-1} \right)^{\frac{1}{2}} f_N \to 0, \quad \text{as} \quad N \to \infty.
\]

Hence, the result of part (c) is also satisfied. \(\square\)
Lemma 2. Let $A \subset \Omega_N^*$. Let $C_N$ denote the cardinality of $A$. If there exist sequences $f_N > 0$ and $d_N \geq 0$ satisfying (a), $0 < C_N < \infty$ and $f_N^2 / \log C_n \to \infty$ as $N \to \infty$, (b). If $y, v \in \Omega_N$ and $\|y - M_N^x\|_{\sigma(A)} \leq f_N$, then there exists $n < \infty$ such that for all $N > n$, $\|U_N(y, v)\|_{\sigma(A)} \leq d_N\|y - M_N^x\|_{\sigma(A)}\|v\|_{\sigma(A)}$, (c). $d_N f_N \to 0$ as $N \to \infty$. Then $\Pr(\|R_N\|_{\sigma(A)} < \frac{1}{2} f_N) \to 1$ as $N \to \infty$.

Proof. Denote $A = \{g_k : k = 1, \ldots, C_N\}$ and let $w_k \in \Omega_N$ be defined for $k = 1, \ldots, C_N$ by $g_k(x) = [w_k, x]_{\sigma}, x \in \Omega_N$. Let $W_k = \|w_k\|_{\sigma}^{-1} \sum_{i=1}^N \sum_{j \in S_N(i)} w_{ijk}(Y_{ij} - E_{ij})$ for $k = 1, \ldots, C_N$ so that $\|R_N\|_{\sigma(A)} = \max_{k=1,\ldots,C_N} |W_k|$. We consider the log moment generating function of $W_k$, denoted as $\log G_k(t)$. Write $w'_k = w_k / \|w_k\|_{\sigma}, k = 1, \ldots, C_N$, for simplicity, and we have

$$\log G_k(t) = \log \mathbb{E}[e^{tW_k}]$$

$$= \log \mathbb{E}\left[\exp\left\{\frac{t}{\|w_k\|_{\sigma}} \sum_{i=1}^N \sum_{j \in S_N(i)} w_{ijk}(Y_{ij} - E_{ij})\right\}\right]$$

$$= -t \sum_{i=1}^N \sum_{j \in S_N(i)} w'_{ijk} E_{ij} + \log \prod_{i=1}^N \prod_{j \in S_N(i)} \mathbb{E}\left\{\exp(tw'_{ijk} Y_{ij})\right\}, \text{ by independence}$$

$$= -t \sum_{i=1}^N \sum_{j \in S_N(i)} w'_{ijk} E_{ij} + \sum_{i=1}^N \sum_{j \in S_N(i)} \log \mathbb{E}\left\{\exp(tw'_{ijk} Y_{ij})\right\}$$

$$= \sum_{i=1}^N \sum_{j \in S_N(i)} \left[\log\left\{1 + \exp(m_{ij}^*)\right\}^{-1} - \log\left\{1 + \exp(tw'_{ijk} + m_{ij}^*)\right\}^{-1} - tw'_{ijk} E_{ij}\right]$$

$$= \sum_{i=1}^N \sum_{j \in S_N(i)} \left[\log\{h(m_{ij}^*)\} - \log\{h(tw'_{ijk} + m_{ij}^*)\} - tw'_{ijk} E_{ij}\right], \quad (64)$$

where we have denoted $h(x) = \{1 + \exp(x)\}^{-1}$. We apply Taylor expansion to $\log\{h(tw'_{ijk} + m_{ij}^*)\}$ with respect to $m_{ij}^*$. For some $t' = \alpha t$ with $0 < \alpha < 1$, we have

$$\log\{h(tw'_{ijk} + m_{ij}^*)\} = \log h(m_{ij}^*) - E_{ij} tw'_{ijk} - \frac{t'^2}{2} w'_{ijk}^2 \sigma^2 (m_{ij}^* + t'w'_{ijk}).$$

Substitute into (64),

$$\log G_k(t) = \frac{t'^2}{2} \sum_{i=1}^N \sum_{j \in S_N(i)} w'_{ijk}^2 \sigma^2 (m_{ij}^* + t'w'_{ijk}), \quad |t| \leq f_N.$$
By Bahadur (1971),
\[
pr\left(W_k \geq \frac{1}{2} f_N\right) \leq \exp\left(-\frac{1}{4} f_N^2\right) G_k\left(\frac{1}{2} f_N\right), \quad k = 1, \ldots, C_N,
\]
and
\[
pr\left(-W_k \geq \frac{1}{2} f_N\right) \leq \exp\left(-\frac{1}{4} f_N^2\right) G_k\left(-\frac{1}{2} f_N\right), \quad k = 1, \ldots, C_N.
\]
Furthermore note that,
\[
\log G_k\left(\frac{1}{2} f_N\right), \log G_k\left(-\frac{1}{2} f_N\right) \leq \frac{1}{8} f_N^2\left(1 + \frac{d_N f_N}{2}\right) \quad k = 1, \ldots, C_N.
\]
Applying the Bonferroni inequality,
\[
pr\left\{\|R_N\|_\sigma(A) \geq \frac{1}{2} f_N\right\} \leq 2C_N \exp\left\{-\frac{1}{8} f_N^2\left(1 - \frac{d_N f_N}{2}\right)\right\}
\]
\[
= 2 \exp\left\{\log C_N - \frac{1}{8} f_N^2\left(1 - \frac{d_N f_N}{2}\right)\right\}
\]
\[
\to 0 \quad \text{as} \quad N \to \infty,
\]
where the last step follows from the assumption \(f_N^2 / \log C_N \to \infty\) as \(N \to \infty\). Hence the result of the lemma follows. \qed

**Lemma 3.** Assume Conditions 1, 2 and 3 hold. Let \(A \subset \Omega_N^\ast\). If there exist sequences \(f_N > 0\) and \(d_N \geq 0\) satisfying (a). \(pr\{\|R_N\|_\sigma(A) < \frac{1}{2} f_N\} \to 1\) as \(N \to \infty\), (b). If \(y, v \in \Omega_N\) and \(\|y - M_N^\ast\|_\sigma(A) \leq f_N\), then there exists \(n < \infty\) such that for all \(N > n\), \(\|U_N(y, v)\|_\sigma(A) \leq d_N\|y - M_N^\ast\|_\sigma(A)\|v\|_\sigma(A)\), (c). \(d_N f_N \to 0\) as \(N \to \infty\). Then, as \(N \to \infty\), with probability approaching 1 that,
\[
\left|\frac{\|\hat{M}_N - M_N^\ast\|_\sigma(A)}{\|R_N\|_\sigma(A)} - 1\right| \leq d_N^{\frac{1}{2}} \to 0 \quad \text{and} \quad \|\hat{M}_N - M_N^\ast - R_N\|_\sigma(A) \leq d_N\|R_N\|_\sigma^2(A).
\]

**Proof.** Write \(z_N = \|R_N\|_\sigma(A)\) for simplicity. Consider a sequence \(\{h_{Nk} : k = 0, 1, \ldots\}\), with \(h_{N0} = 0\)
and $h_{N(k+1)} = z_N + d_N h_{Nk}^2/2$ for $k = 0, 1, 2, \ldots$. Define another sequence

$$l_N = \frac{2z_N}{1 + (1 - 2z_N d_N)^{1/2}}.$$  

By Kantorovich and Akilov (1964, pages 695-711), if $z_N < 1/2$ and $z_N d_N < 1/2$ (which hold with probability tending to 1 by (a), (b) and (c)), it follows

$$\|t_{Nk} - \hat{M}_N\|_\sigma(A) \leq l_N - h_{Nk}, \quad k = 0, 1, 2, \ldots,$$  

(65)

where $\{t_{Nk} : k = 0, 1, \ldots\}$ is the sequence constructed in the proof of Theorem 1. When $k = 0$, (65) implies $\|M_N^* - \hat{M}_N\|_\sigma(A) \leq l_N$. When $k = 1$, (65) implies

$$\|M_N^* + R_N - \hat{M}_N\|_\sigma(A) \leq l_N - z_N.$$  

(66)

It follows that $\|M_N^* - \hat{M}_N\|_\sigma(A) - \|R_N\|_\sigma(A) \leq l_N - z_N$, where

$$l_N - z_N = \frac{z_N(1 - (1 - 2z_N d_N)^{1/2})}{1 + (1 - 2z_N d_N)^{1/2}}.$$

If we view $x = z_N d_N$ and $f(x) = \{1 - (1 - 2x)\}^{1/2}/\{1 + (1 - 2x)\}^{1/2}$. We note $f(0) = 0, f(1/2) = 1$ and $f'(0) = 1/4 < 1$ and $f''(x) > 0$ for all $x < 1/2$. Therefore, $f(x) < x$ for all $x < 1/2$. Hence, whenever $d_N z_N < 1/2$, we must have $l_N - z_N \leq d_N z_N^2$. We know that with probability tending to 1 that $d_N z_N < 1/2$. Hence the second part of the lemma follows from (66). Also as $N \to \infty$, with probability approaching 1 that,

$$\left|\|\hat{M}_N - M_N^*\|_\sigma(A) - \|R_N\|_\sigma(A)\right|^2 \leq d_N \|R_N\|_\sigma^2(A).$$  

(67)

Re-write (67), the result of the first part of the lemma then follows.

\[ \square \]

**Lemma 4.** Assume Conditions 1, 2 and 3 hold and $\sum_{i=1}^{N} \theta_i = 0$, the asymptotic variance of the maximum likelihood estimator of $m_{ij}^*$, $\text{var}(\hat{m}_{ij})$, for any $i = 1, \ldots, N$ and $j = 1, \ldots, J$, takes the
form,

\[ \text{var}(\hat{m}_{ij}) = (\sigma_{i+}^2)^{-1} + (\sigma_{+j}^2)^{-1} + O(N_s^{-1}J_s^{-1}) \quad \text{as} \quad N \to \infty. \]

**Proof.** If \( z_{ij} = 1 \), then we can simply use a linear function \( f_{ij} \) with \( f_{ij}(x) = x_{ij} \). We apply \( \|d'(f_{ij})\|_{\sigma}^2 \) to approximate \( \sigma^2(f_{ij}) \). With \( w_{i+} = 1, w_{k+} = 0 \), for all \( k = 1, ..., i-1, i+1, ..., N, w_{+j} = 1, w_{+l} = 0 \) for all \( l = 1, ..., j-1, j+1, ..., J \) and \( w_{++} = 1 \). We obtain

\[ \|d'(f_{ij})\|_{\sigma}^2 = (\sigma_{i+}^2)^{-1} + (\sigma_{+j}^2)^{-1} + O\left(N_s^{-1}J_s^{-1}\right) \quad \text{as} \quad N \to \infty. \]

If \( z_{ij} = 0 \), then we can apply Condition 3, there must exist \( 1 \leq i_1, i_2, ..., i_k \leq N \) and \( 1 \leq j_1, j_2, ..., j_k \leq J \) such that \( z_{ij_1} = z_{i_1j_1} = z_{i_1j_2} = z_{i_2j_2} = ... = z_{i_kj_k} = z_{i_kj} = 1 \). Consider a linear function \( g_2 \) defined as

\[
g_{ij}(x) = x_{ij_1} - x_{i_1j_1} + x_{i_1j_2} - x_{i_2j_2} + ... + x_{i_{k-1}j_k} - x_{i_kj_k} + x_{i_kj}
\]

\[ = \theta_i - \beta_j. \]

In this case, similarly we have \( w_{i+} = 1, w_{k+} = 0 \), for all \( k = 1, ..., i-1, i+1, ..., N, w_{+j} = 1, w_{+l} = 0 \) for all \( l = 1, ..., j-1, j+1, ..., J \) and \( w_{++} = 1 \). Note these values are exactly the same as those of \( g_1 \). Therefore,

\[ \|d'(g_{ij})\|_{\sigma}^2 = (\sigma_{i+}^2)^{-1} + (\sigma_{+j}^2)^{-1} + O\left(N_s^{-1}J_s^{-1}\right) \quad \text{as} \quad N \to \infty. \]

In both cases, \( \|d''\|_{\sigma}^2 \) has a small order. To see this, note that in both cases above,

\[
l_p = \begin{cases} 
-\sigma_{pj}^2(\sigma_{+j}^2)^{-1} + \sigma_{p+}^2(\sigma_{++}^2)^{-1} & \text{if} \quad z_{pj} = 1 \\
\sigma_{p+}^2(\sigma_{++}^2)^{-1} & \text{if} \quad z_{pj} = 0
\end{cases}
\]

\[ = O\left(N_s^{-1}\right) \quad \text{as} \quad N \to \infty, \quad p = 1, ..., N. \]
\[ v_q = \begin{cases} 
-\sigma^2_{i+}(\sigma^2_{i+})^{-1} + \sigma^2_{p+}(\sigma^2_{p+})^{-1} & \text{if } z_{iq} = 1 \\
\sigma^2_{p+}(\sigma^2_{p+})^{-1} & \text{if } z_{iq} = 0 
\end{cases} \]

\[ = O(J_*^{-1}) \quad \text{as } N \to \infty, \quad q = 1, \ldots, J. \]

It follows that

\[
\|d''(f_{ij})\|_\sigma^2 = \|d''(g_{ij})\|_\sigma^2 \leq \gamma_N^{-2} \left\{ \sum_{p=1}^{N} l^2_p(\sigma^2_{p+})^{-1} + \sum_{q=1}^{J} v^2_q(\sigma^2_{p+})^{-1} \right\}^2 
\]

\[ = O(N_*^{-1}J_*^{-2}) \quad \text{as } N \to \infty, \]

where the last line follows from Condition 1(a). Since for any \( g \in \Omega_N^* \),

\[
(\|d'(g)\|_\sigma - \|d''(g)\|_\sigma)^2 \leq \sigma^2(g) \leq (\|d'(g)\|_\sigma + \|d''(g)\|_\sigma)^2,
\]

it follows \( \text{var}(\hat{m}_{ij}) = (\sigma^2_{i+})^{-1} + (\sigma^2_{p+})^{-1} + O(N_*^{-1}J_*^{-1}) \) as \( N \to \infty. \)

**Lemma 5.** Assume Conditions 1, 2 and 3 hold. If \( A_\beta = \{g_j : j = 1, \ldots, J\} \) such that \( g_j \in \Omega_N^* \) and \( g_j(x) = \beta_j \) for \( x \in \Omega_N \). Let \( C_N = |A_\beta| = J \) be the cardinality of \( A_\beta \). For any positive sequence \( f_N \) such that \( f_N^2/\log J \to \infty \) and \( f_N^2N_*^{-1/2} \to 0 \) as \( N \to \infty \), there exists a sequence \( d_N \geq 0 \) satisfying the followings.

(a). If \( y, v \in \Omega_N \) and \( \|y - M_N^\sigma(A_\beta)\|_\sigma \leq f_N \), then there exists \( n < \infty \) such that for all \( N > n \),

\[
\|U_N(y, v)\|_\sigma(A_\beta) \leq d_N\|y - M_N\|_\sigma(A_\beta)\|v\|_\sigma(A_\beta).
\]

(b). \( d_Nf_N^2 \to 0 \) as \( N \to \infty \).

**Proof.** First we note since we have assumed \( N > J \), we must have \( \log J \ll N_*^{1/2} \) by Condition 1(a), so the rate requirements for \( f_N \) is valid. To find a valid \( d_N \), we seek to upper bound \( \|U_N(y, v)\|_\sigma(A_\beta) \) and then show that \( d_Nf_N \to 0 \) as \( N \to \infty \) for all \( f_N \) satisfying the rate requirements \( f_N^2/\log J \to \infty \) and \( f_N^2N_*^{-1/2} \to 0 \) as \( N \to \infty \). For any \( y, v \in \Omega_N \), by the definition of \( \| \cdot \|_\sigma(A_\beta) \), we have

\[
\|U_N(y, v)\|_\sigma(A_\beta) = \max_{g_j \in A_\beta} |g_j \{U_N(y, v)\}| / \sigma(g_j).
\]
First note that by Lemma 6, \( \sigma^2(g_j) = (\sigma_{ij}^2)^{-1} + O\{N_*J_*\}^{-1} \) for any \( g_j \in A_\beta \). Therefore, there exist positive \( 0 < c_1, c_2 < \infty \) such that for all \( N > n, c_1^{-1}N_*^{-1/2} < \sigma(g_j) < c_2^{-1}N_*^{-1/2} \), for all \( g_j \in A_\beta \). So we just need to find an upper bound for \( |g_j\{U_N(y, v)\}| \) that holds for all \( g_j \in A_\beta \).

Consider

\[
|g_j\{U_N(y, v)\}| = \left| \sum_{i'=1}^N \sum_{j'\in S_j(i')} d_{i'j'}(g_j)\{\sigma^2(y_{i'j'}) - \sigma_{i'j'}^2\} v_{i'j'} \right| \\
\leq \sum_{i'=1}^N \sum_{j'\in S_j(i')} |d_{i'j'}(g_j)| \cdot |\sigma^2(y_{i'j'}) - \sigma_{i'j'}^2| \cdot |v_{i'j'}|.
\]

Note \( 0 \leq \sigma^2(y_{ij}), \sigma_{ij}^2 \leq 1 \), so \( |\sigma^2(y_{ij}) - \sigma_{ij}^2| \leq 1 \). It can be implied that there exists some positive \( c_3 < \infty \) such that \( |\sigma^2(y_{ij}) - \sigma_{ij}^2| \leq c_3|g_j(y - M_N^*)| \). Again, by the definition of \( \| \cdot \|_\sigma(A_\beta) \), we have

\[
|g_j(y - M_N^*)| \leq \|y - M_N^*\|_\sigma(A_\beta)\sigma(g_j).
\]

Therefore, for all \( i = 1, \ldots, N, j = 1, \ldots, J, z_{ij} = 1 \),

\[
|\sigma^2(y_{ij}) - \sigma_{ij}^2| \leq c_2^{-1}c_3N_*^{-1/2}\|y - M_N^*\|_\sigma(A_\beta).
\]

On the other hand, using a similar strategy, we can show that there exists a positive \( c_4 < \infty \) such that for all \( i = 1, \ldots, N, j = 1, \ldots, J, z_{ij} = 1 \),

\[
|v_{ij}| \leq c_2^{-1}c_4N_*^{-1/2}\|v\|_\sigma(A_\beta).
\]

Further note that

\[
\sum_{i'=1}^N \sum_{j'\in S_j(i')} |d_{i'j'}(g_j)| \leq \sum_{i'=1}^N \sum_{j'\in S_j(i')} |d'_{i'j'}(g_j)| + \sum_{i'=1}^N \sum_{j'\in S_j(i')} |d''_{i'j'}(g_j)|.
\]

By definition, we know \( d'_{i'j'} = (\sigma_{i'j'}^2)^{-1}w_{i'j'} + (\sigma_{i'j'}^2)^{-1}w_{i'j'} - (\sigma_{i'j'}^2)^{-1}w_{i'j'} \). For any \( g_j \in A_\beta \), \( w_{i'j'} = -1/N, \) for \( i' = 1, \ldots, N, w_{i'j'} = -1 \) if \( j' = j \) and \( w_{i'j'} = 0 \) if \( j' \neq j \), \( w_{i'j'} = -1 \). Hence,

\[
d'_{i'j'}(g_j) = \begin{cases} 
-\frac{1}{N}(\sigma_{i'j'}^2)^{-1} - (\sigma_{i'j'}^2)^{-1} + (\sigma_{i'j'}^2)^{-1} & \text{if } j' = j \\
-\frac{1}{N}(\sigma_{i'j'}^2)^{-1} + (\sigma_{i'j'}^2)^{-1} & \text{if } j' \neq j.
\end{cases}
\]
It follows
\[
\sum_{i'=1}^{N} \sum_{j' \in S_j(i')} |d'_{i',j'}(g_j)| \leq \frac{J^*}{N} \sum_{i'=1}^{N} (\sigma^2_{i'j'})^{-1} + N^* (\sigma_{i'j'}^2)^{-1} + \sum_{i'=1}^{N} \sum_{j' \in S_j(i')} (\sigma^2_{i'j'})^{-1} \leq c_5,
\]
for some positive \( c_5 < \infty \). On the other hand,
\[
\sum_{i'=1}^{N} \sum_{j' \in S_j(i')} |d''_{i',j'}(g_j)| \leq (N^* J^*)^{1/2} \|d''(g_j)\|_\sigma \leq c_6,
\]
for some positive \( c_6 < \infty \). The last step follows from Condition 1(c) and Lemma 6 which implies that \( \|d''(g_j)\|^2_\sigma = O \left( N^{-1} J^{-1}_* \right) \). Overall,
\[
\|U_N(y,v)\|_\sigma(A_\beta) = \max_{g_j \in A_\beta} |g_j\{U_N(y,v)\}|/\sigma(g_j) \\
\leq \max_{g_j \in A_\beta} |g_j\{U_N(y,v)\}| \cdot \max_{g_j \in A_\beta} \{\sigma^{-1}(g_j)\} \\
\leq c_1 c_2^{-2} c_3 c_4 (c_5 + c_6) N_*^{-1/2} \|y - M_N\|_\sigma(A_\beta)/\sigma(A_\beta).
\]
Note that by taking \( d_N = c_1 c_2^{-2} c_3 c_4 (c_5 + c_6) N_*^{-1/2} \), part (a) of the lemma follows. Furthermore, by the rate requirement of \( f_N \), for any positive sequence \( f_N \) such that \( \log J \ll f_N^2 \ll N_*^{1/2} \), it can be seen easily that \( d_N f_N^2 \rightarrow 0 \) as \( N \rightarrow \infty \). Therefore, part (b) of the lemma follows.

**Lemma 6.** Assume Conditions 1, 2 and 3 hold and \( \sum_{i=1}^{N} \theta_i = 0 \). The asymptotic variance of the maximum likelihood estimator of an individual column parameter, \( \var(\hat{\beta}_j) \), asymptotically attains the oracle variance \( (\sigma^2_{i'j'})^{-1} \) in the sense that
\[
\var(\hat{\beta}_j) = (\sigma^2_{i'j'})^{-1} + O(N^{-1} J^{-1}_*) \quad \text{as} \quad N \rightarrow \infty.
\] (68)

**Proof.** We seek to construct a linear function \( g_j \in \Omega_N^* \) such that \( g_j(x) = \beta_j \) so that we can use \( \|d'(g_j)\|^2_\sigma \) defined in Lemma 13 to approximate \( \var(\hat{\beta}_j) \). To construct such a \( g_j \), we may want to include all \( x_{ij} \), \( i = 1, ..., N \), in \( g_j \) so that we can apply the constraint \( \sum_{i=1}^{N} \theta_i = 0 \) to solve for \( \beta_j \). For \( i \in S_N(j) \), we use \( x_{ij} = \theta_i - \beta_j \) directly. For each \( i \in S_N(j) \), by Condition 3, there must exist \( 1 \leq i_1, i_2, ..., i_k \leq N \) and \( 1 \leq j_1, j_2, ..., j_k \leq J \) such that \( z_{i,j_1} = z_{i_1,j_1} = z_{i_1,j_2} = z_{i_2,j_2} = ... = z_{i_k,j_k} = \ldots \)
\[ z_{i,k,j,k} = z_{i,k,j} = 1, \text{ with} \]
\[
\begin{align*}
x_{i,j_1} - x_{i_1,j_1} + x_{i_1,j_2} - x_{i_2,j_2} + \ldots - x_{i,k,j_k} + x_{i,k,j} \\
= (\theta_i - \beta_{j_1}) - (\theta_{i_1} - \beta_{j_1}) + (\theta_{i_1} - \beta_{j_2}) - (\theta_{i_2} - \beta_{j_2}) + \ldots - (\theta_{i,k} - \beta_{j_k}) + (\theta_{i,k} - \beta_j) \\
= \theta_i - \beta_j.
\end{align*}
\]

Therefore, we can construct \( g \) to be
\[
g_j(x) = -\frac{1}{N} \left\{ \sum_{i \in S_N(j)} x_{ij} \\
+ \sum_{i \in S_{N_2}(j)} \left( x_{i,j_1} - x_{i_1,j_1} + x_{i_1,j_2} - x_{i_2,j_2} + \ldots - x_{i,k,j_k} + x_{i,k,j} \right) \right\}
= -\frac{1}{N} \left\{ \left( \sum_{i=1}^N \theta_i \right) - N\beta_j \right\}
= \beta_j.
\]

Use \( \|d'(g_j)\|_\sigma^2 \) from Lemma 13 to approximate \( \sigma^2(g_j) \), with \( w_{i+} = -1/N \), for all \( i = 1, \ldots, N \), \( w_{+j} = -1 \), \( w_{+l} = 0 \) for all \( l = 1, \ldots, j - 1, j + 1, \ldots, J \) and \( w_{++} = -1 \). It follows
\[
\|d'(g_j)\|_\sigma^2 = (\sigma^2_{i,j})^{-1} + \frac{1}{N^2} \sum_{i=1}^N (\sigma^2_{i,+})^{-1} + \frac{2}{N} \sum_{i \in S_N(j)} \sigma^2_{ij} (\sigma^2_{i,+})^{-1} (\sigma^2_{j,+})^{-1} - 3(\sigma^2_{++})^{-1}
= (\sigma^2_{+j})^{-1} + O \left( N_*^{-1} J_*^{-1} \right) \quad \text{as} \quad N \to \infty.
\]

To see whether \( \|d''(g_j)\|_\sigma^2 \) is a good approximation for \( \sigma^2(g_j) \), we need to evaluate the order of \( \|d''(g_j)\|_\sigma^2 \) from Lemma 14. Note
\[
l_i = \begin{cases} 
\sigma^2_{ij} (\sigma^2_{i,j})^{-1} - \sigma^2_{i,+} (\sigma^2_{j,+})^{-1} & \text{if} \ z_{ij} = 1 \\
-\sigma^2_{i,+} (\sigma^2_{j,+})^{-1} & \text{if} \ z_{ij} = 0
\end{cases}
= O \left( N_*^{-1} \right) \quad \text{as} \quad N \to \infty, \quad i = 1, \ldots, N.
\[ v_q = \frac{1}{N} \sum_{i \in S_N(q)} \sigma_i^2 (\sigma_i^2)^{-1} - \sigma_{ij}^2 (\sigma_{ij}^2)^{-1} \]

\[ = O \left( J_*^{-1} \right) \quad \text{as} \quad N \to \infty, \quad q = 1, \ldots, J. \]

Applying Lemma 14, we have

\[ \|d''(g_j)\|_2^2 \leq \gamma_N^{-2} \left\{ \sum_{i=1}^{N} l_i^2 (\sigma_i^2)^{-1} + \sum_{q=1}^{J} v_q^2 (\sigma_{iq}^2)^{-1} \right\}^2 \]

\[ = O \left( N_*^{-1} J_*^{-2} \right) \quad \text{as} \quad N \to \infty, \]

where the last line follows from Condition 1(a). Since

\[ (\|d'(g_j)\|_\sigma - \|d''(g_j)\|_\sigma)^2 \leq \sigma^2(g_j) \leq (\|d'(g_j)\|_\sigma + \|d''(g_j)\|_\sigma)^2, \]

It follows that \( \text{var}(\hat{\beta}_j) = (\sigma_{ij}^2)^{-1} + O \left( N_*^{-1} J_*^{-1} \right) \) as \( N \to \infty. \)

\[ \square \]

**Lemma 7.** Let \( a_N \) and \( c_N \) be positive sequences. As \( N \to \infty \), suppose that \( a_N = O(b_N) \), for any sequence \( b_N \) satisfying \( b_N/c_N \to \infty \). Then \( a_N = O(c_N) \) as \( N \to \infty \).

**Proof.** We prove the lemma result by a contradiction argument. Suppose the lemma result does not hold. Then there exists a subsequence \( N_n \), such that

\[ L_{N_n} = \frac{a_{N_n}}{c_{N_n}} \to \infty. \]

We define a sequence

\[ b_N = \begin{cases} N c_N & \text{if } N \notin \{N_1, N_2, \ldots\}, \\ \frac{1}{L_{N_n}^2} c_{N_n} & \text{if } N = N_n. \end{cases} \]

Then \( b_N/c_N \to \infty \), but \( a_N = O(b_N) \) does not hold. Contradiction. Hence, the lemma result holds.

\[ \square \]

**Lemma 8.** Assume Conditions 1, 2 and 3 hold. If \( A_\theta = \{g_i : i = 1, \ldots, N\} \) such that \( g_i \in \Omega_N^* \)
and \(g_i(x) = \theta_i\) for \(x \in \Omega_N\). Let \(C_N = |A_\theta| = N\) be the cardinality of \(A_\theta\). Then for any positive sequence \(f_N\) such that \(f_N^2 / \log N \to \infty\) and \(J_*^{-1}f_N^2 \to 0\) as \(N \to \infty\), there exists a sequence \(d_N \geq 0\) satisfying the followings.

(a) If \(y, v \in \Omega_N\) and \(\|y - M_N^y\|_\sigma(A_\theta) \leq f_N\), then there exists \(n < \infty\) such that for all \(N > n\), \(\|U_N(y, v)\|_\sigma(A_\theta) \leq d_N\|y - M_N^y\|_\sigma(A_\theta)\|v\|_\sigma(A_\theta)\).

(b) \(d_Nf_N \to 0\) as \(N \to \infty\).

Proof. We first note that from Condition 1(b), \(\log N \ll J_*\) as \(N \to \infty\). Therefore, the rate requirements for the sequence \(f_N, f_N^2 / \log N \to \infty\) and \(J_*^{-1}f_N^2 \to 0\) as \(N \to \infty\), are valid. Now we seek to upper bound \(\|U_N(y, v)\|_\sigma(A_\theta)\) to find a sequence \(d_N\) and then show that \(d_Nf_N \to 0\) for any \(f_N\) satisfying \(f_N^2 / \log N \to \infty\) and \(J_*^{-1}f_N^2 \to 0\) as \(N \to \infty\). For any \(y, v \in \Omega_N\), by the definition of \(\| \cdot \|_\sigma(A_\theta)\),

\[
\|U_N(y, v)\|_\sigma(A_\theta) = \max_{g_i \in A_\theta} |g_i\{U_N(y, v)\}|/\sigma(g_i).
\]

Note that by Lemma 9, we know that \(\sigma^2(g_i) = (\sigma^2_{i+1})^{-1} + O\{N_*^{-1}J_*^{-1}\}\) for any \(g_i \in A_\theta\). Hence, there exist positive \(0 < \gamma_1, \gamma_2 < \infty\) such that for any \(i = 1, \ldots, N\),

\[
\gamma_1^{-1}J_*^{-1/2} < \sigma(g_i) < \gamma_2^{-1}J_*^{-1/2}.
\]

So we just need to find an upper bound for \(|g_i\{U_N(y, v)\}|\) that holds for all \(g_i \in A_\theta\). For any \(g_i \in A_\theta\), we have

\[
|g_i\{U_N(y, v)\}| = \left| \sum_{i' = 1}^N \sum_{j' \in S_{j'}(i')} d_{i'j'}(g_i)\{\sigma^2(y_{i'j'}) - \sigma^2_{i'j'}\}v_{i'j'}\right|
\]

\[
\leq \sum_{i' = 1}^N \sum_{j' \in S_{j'}(i')} |d_{i'j'}(g_i)| \cdot |\sigma^2(y_{i'j'}) - \sigma^2_{i'j'}| \cdot |v_{i'j'}|.
\]

Since \(\sigma^2(y_{ij}), \sigma^2_{ij} < 1\), so \(|\sigma^2(y_{ij}) - \sigma^2_{ij}| \leq 1\). It can be implied that there exists a positive \(\gamma_3 < \infty\) such that \(|\sigma^2(y_{ij}) - \sigma^2_{ij}| \leq \gamma_3|g_i(y - M_N^y)|\). From the definition of \(\| \cdot \|_\sigma(A_\theta)\), \(|g_i(y - M_N^y)| \leq \|y - M_N^y\|_\sigma(A_\theta)\) for any \(g_i \in A_\theta\). Then it follows that for any \(i = 1, \ldots, N, j = 1, \ldots, J, z_{ij} = 1\),

\[
|\sigma^2(y_{ij}) - \sigma^2_{ij}| \leq \gamma_2^{-1}\gamma_3J_*^{-1/2}\|y - M_N^y\|_\sigma(A_\theta).
\]
Using a similar strategy, we can also show that there exists a positive \( \gamma_4 < \infty \) such that for any \( i = 1, \ldots, N, j = 1, \ldots, J, z_{ij} = 1 \),

\[
|v_{ij}| \leq \gamma_2^{-1} \gamma_4 J_{**}^{-1/2} \|v\|_\sigma(A_\theta).
\]

Similarly, we have

\[
\sum_{i' = 1}^{N} \sum_{j' \in S_j(i')} |d''_{i'j'}(g_i)| \leq \sum_{i' = 1}^{N} \sum_{j' \in S_j(i')} |d''_{i'j'}(g_i)| + \sum_{i' = 1}^{N} \sum_{j' \in S_j(i')} |d''_{i'j'}(g_i)|.
\]

By definition, we know \( d''_{i'j'} = (\sigma_{i'j'}^2)^{-1} w_{i'j'} + (\sigma_{i'j'}^2)^{-1} w_{jj'} - (\sigma_{j+j}^2)^{-1} w_{++}. \) For any \( g_i \in A_\theta \), \( w_{i'j'} = 1 - 1/N \), if \( i' = i \), and \( w_{i'j'} = -1/N \) for \( i' \neq i \), \( w_{jj'} = 0 \) for all \( j' = 1, \ldots, J \) and \( w_{++} = 0 \). Hence,

\[
d''_{i'j'}(g_i) = \begin{cases} 
(1 - \frac{1}{N})(\sigma_{i'j'}^2)^{-1} & \text{if } i' = i \\
-\frac{1}{N}(\sigma_{j+j}^2)^{-1} & \text{if } i' \neq i.
\end{cases}
\]

It follows

\[
\sum_{i' = 1}^{N} \sum_{j' \in S_j(i')} |d''_{i'j'}(g_i)| = \sum_{j' \in S_j(i')} \left(1 - \frac{1}{N}\right)(\sigma_{i'j'}^2)^{-1} + \sum_{i' = 1, i' \neq i}^{N} \sum_{j' \in S_j(i')} \frac{1}{N}(\sigma_{j+j}^2)^{-1} \leq \gamma_5,
\]

for some positive \( \gamma_5 < \infty \). On the other hand,

\[
\sum_{i' = 1}^{N} \sum_{j' \in S_j(i')} |d''_{i'j'}(g_j)| \leq (N^* J^*)^{1/2} \|d''(g_j)\|_\sigma \leq \gamma_6,
\]

for some positive \( \gamma_6 < \infty \). The last step follows from Condition 1(c) and Lemma 9 which implies that \( \|d''(g_j)\|_\sigma^2 = O \left( N_*^{-1} J_*^{-1} \right) \). Overall,

\[
\|U_N(y, v)\|_\sigma(A_\theta) = \max_{g_i \in A_\theta} \|g_i \{U_N(y, v)\}\|/\sigma(g_i)
\leq \max_{g_i \in A_\theta} \|g_i \{U_N(y, v)\}\| \cdot \max_{g_i \in A_\theta} \{\sigma^{-1}(g_i)\}
\leq \gamma_1 \gamma_2^{-2} \gamma_3 \gamma_4 (\gamma_5 + \gamma_6) J_* \gamma_*^{-1} \|y - M_N^*\|_\sigma(A_\theta) \|v\|_\sigma(A_\theta).
\]

65
So we can set \( d_N = \gamma_1 \gamma_2 \gamma_3 \gamma_4 (\gamma_5 + \gamma_6) J_*^{1/2} \). Furthermore, by the rate requirement of \( f_N \), for any positive sequence \( f_N \) such that \((\log N)^{1/2} \ll f_N \ll J_*^{1/2} \), we must have \( d_N f_N \to 0 \) as \( N \to \infty \). Therefore, both part (a) and part (b) of the lemma are satisfied. \( \square \)

**Lemma 9.** Assume Conditions 1, 2 and 3 hold and \( \sum_{i=1}^{N} \theta_i = 0 \), the asymptotic variance of an individual row parameter, \( \text{var}(\hat{\theta}_i) \), asymptotically attains oracle variance \((\sigma_{i+}^2)^{-1}\) in the sense that

\[
\text{var}(\hat{\theta}_i) = (\sigma_{i+}^2)^{-1} + O \left( N_*^{-1} J_*^{-1} \right) \quad \text{as} \quad N \to \infty.
\]  

(69)

**Proof.** We seek to construct a linear function \( g_i \in \Omega_N^* \) such that \( g_i(x) = \theta_i \) so that we can use \( \|d'(g_i)\|_\sigma^2 \) in Lemma 13 to approximate \( \text{var}(\hat{\theta}_i) \). Fix some \( j \in S_J(i) \), i.e. \( z_{ij} = 1 \), since Condition 3 holds, we can use the linear function \( g_j \) constructed in the proof of Theorem 1 to represent \( \beta_j \), i.e. \( g_j(x) = \beta_j \). Hence, \( g_i \) can easily be constructed with

\[
g_i(x) = \frac{1}{|S_J(i)|} \sum_{j \in S_J(i)} \{ x_{ij} + g_j(x) \}
\]

\[
= \frac{1}{|S_J(i)|} \sum_{j \in S_J(i)} \left[ x_{ij} - \frac{1}{N} \sum_{i' \in S_N(j)} x_{i'j} \right]
+ \sum_{i' \in S_{N_*}(j)} \left( x_{i',j'1} - x_{i'1,j'1} + x_{i'1,j'2} - x_{i'2,j'2} + \ldots - x_{i'k,j'k} + x_{i'k,j} \right)
\]

\[
= \theta_i.
\]

We use \( \|d'(g_i)\|_\sigma^2 \) from Lemma 13 to approximate \( \sigma^2(g_i) \), with \( w_{i+} = 1 - N^{-1}, w_{k+} = -N^{-1} \), for all \( k = 1, \ldots, i - 1, i + 1, \ldots, N \), \( w_{++} = 0 \), for all \( j = 1, \ldots, J \), \( w_{++} = 0 \), we obtain

\[
\|d'(g_i)\|_\sigma^2 = \left( 1 - \frac{1}{N} \right)^2 (\sigma_{i+}^2)^{-1} + \frac{1}{N^2} \sum_{k=1, k \neq i}^{N} (\sigma_{k+}^2)^{-1}
\]

\[
= (\sigma_{i+}^2)^{-1} + O \left( N_*^{-1} J_*^{-1} \right) \quad \text{as} \quad N \to \infty.
\]

To see whether \( \|d'(g_i)\|_\sigma^2 \) is a good approximation for \( \sigma^2(g_i) \), we evaluate the order of \( \|d''(g_i)\|_\sigma^2 \).
Note that in this case

\[ l_p = 0, \quad p = 1, \ldots, N. \]

\[ v_q = \begin{cases} 
\frac{1}{N} \sum_{k \in S_N(q), k \neq i} \sigma_k^2 \sigma_k \sigma_i \quad & \text{if } z_{iq} = 1 \\
\frac{1}{N} \sum_{k \in S_N(q)} \sigma_k^2 \sigma_k \sigma_i \quad & \text{if } z_{iq} = 0 
\end{cases} \]

\[ = O \left( J^{-1}_* \right) \quad \text{as } N \to \infty, \quad q = 1, \ldots, J. \]

It follows that

\[
\|d^\prime\prime(g_i)\|_\sigma^2 \leq \gamma_N^{-2} \left\{ \sum_{i=1}^N l_i^2 (\sigma_i^2)^{-1} + \sum_{q=1}^J v_q^2 (\sigma^2_{q+})^{-1} \right\}^2 \\
= O \left( N^{-2} J^{-1}_* \right) \quad \text{as } N \to \infty,
\]

where the last line follows from Condition 1(a). Since

\[
\left( \|d^\prime(g_i)\|_\sigma - \|d^\prime\prime(g_i)\|_\sigma \right)^2 \leq \sigma^2 \left( \|d^\prime(g_i)\|_\sigma + \|d^\prime\prime(g_i)\|_\sigma \right)^2,
\]

it follows that \( \text{var}(\hat{\theta}_i) = (\sigma_i^2)^{-1} + O(N^{-1} J^{-1}_*) \) as \( N \to \infty. \) \( \square \)

**Lemma 10.** Assume Conditions 1, 2 and 3 hold and \( \sum_{i=1}^N \theta_i = 0. \) Consider a linear function \( g : \Omega_N \to \mathbb{R} \) with \( g(M) = \sum_{i=1}^N h_i \theta_i + \sum_{j=1}^J h_j^\prime \beta_j. \) If there exists a positive \( C < \infty \) such that \( \sum_{i=1}^N |h_i| < C \) and \( \sum_{j=1}^J |h_j^\prime| < C, \) then

\[
\sigma^2(g) = \sum_{i=1}^N h_i^2 (\sigma_i^2)^{-1} + \sum_{j=1}^J h_j^\prime (\sigma_j^2)^{-1} + O \left( N^{-1} J^{-1}_* \right) \quad \text{as } N \to \infty.
\]

**Proof.** By Proposition 1, we can reexpress function \( g \) in terms of \( m_{ij} \) for \( i = 1, \ldots, N, j = 1, \ldots, J, z_{ij} = \) 67
1 with \( g(M_N) = \sum_{i=1}^N \sum_{j \in S_j(i)} w_{ij}(g)m_{ij} \). In particular, we have,

\[
w_{i+}(g) = h_i \left( 1 - \frac{1}{N} \right) - \frac{1}{N} \sum_{i'=1, i' \neq i}^N h_{i'} - \frac{1}{N} \sum_{j=1}^J h'_j \quad i = 1, \ldots, N,
\]

\[
w_{+j}(g) = -h'_j, \quad j = 1, \ldots, J,
\]

\[
w_{++}(g) = -\sum_{j=1}^J h'_j.
\]

We apply \( \|d'(g)\|^2_\sigma \) from Lemma 13 to approximate \( \sigma^2(g) \). Note that

\[
\|d'(g)\|^2_\sigma = \sum_{i=1}^N w_{i+}^2(g)(\sigma^2_{i+})^{-1} + \sum_{j=1}^J w_{+j}^2(g)(\sigma^2_{+j})^{-1} + 2 \sum_{i=1}^N \sum_{j \in S_j(i)} \sigma^2_{ij}(\sigma^2_{i+})^{-1}w_{i+}(g)(\sigma^2_{+j})^{-1}w_{+j}(g) - 3(\sigma^2_{++})^{-1}w_{++}^2(g)
\]

\[
= \sum_{i=1}^N h_i^2(\sigma^2_{i+})^{-1} + \sum_{j=1}^J h_j^2(\sigma^2_{+j})^{-1} + O\left( N^{-1} J \right) \quad \text{as} \quad N \to \infty,
\]

where the last step follows from the assumption that \( \sum_{i=1}^N |h_i| < C \) and \( \sum_{j=1}^J |h'_j| < C \). To see whether \( \|d'(g)\|^2_\sigma \) is a good approximation for \( \sigma^2(g) \), we need to evaluate the order of \( \|d''(g)\|^2_\sigma \). Note that for \( i = 1, \ldots, N, \)

\[
l_i = -\sum_{j \in S_j(i)} \sigma^2_{ij}(\sigma^2_{+j})^{-1}w_{+j}(g) + \sigma^2_{i+}(\sigma^2_{++})^{-1}w_{++}(g)
\]

\[
= \sum_{j \in S_j(i)} \sigma^2_{ij}(\sigma^2_{+j})^{-1}h'_j - \sigma^2_{i+}(\sigma^2_{++})^{-1} \sum_{j=1}^J h'_j = O\left( N^{-1} \right) \quad \text{as} \quad N \to \infty, \quad (70)
\]
where the last step follows from $\sum_{j=1}^J |h_j'| < C$. Similarly for $j = 1, ..., J$,

$$v_j = - \sum_{i \in S_N(j)} \sigma_{ij}^2 (\sigma_{ii+}^2)^{-1} w_i(g) + \sigma_{++j}^2 (\sigma_{++}^2)^{-1} w_{++}(g)$$

$$= - \sum_{i \in S_N(j)} \sigma_{ij}^2 (\sigma_{ii+}^2)^{-1} \left\{h_i \left(1 - \frac{1}{N}\right) - \frac{1}{N} \sum_{i' \neq i} h_{i'} - \frac{1}{N} \sum_{j=1}^J h_j' \right\}$$

$$- \sigma_{++j}^2 (\sigma_{++}^2)^{-1} \sum_{j=1}^J h_j'$$

$$= O\left(J_s^{-1}\right) \quad \text{as} \quad N \to \infty,$$

(71)

where the last step follows from $\sum_{j=1}^J |h_j'| < C$ and $\sum_{i=1}^N |h_i| < C$. Hence, we have

$$\|d''(g)\|^2_{\sigma} \leq \gamma_N^{-2} \left\{ \sum_{i=1}^N \|i^2 (\sigma_{ii+}^2)^{-1} + \sum_{j=1}^J v_j^2 (\sigma_{++j}^2)^{-1} \right\}^2$$

$$= O\left(N_s^{-1} J_s^{-2}\right) \quad \text{as} \quad N \to \infty,$$

where the last step can be seen easily from (70), (71) and Condition 1(a) that there exists a constant $k > 0$, such that $N_s \geq k N^{2/3}$ and $J_s \geq k J^{2/3}$. It follows that

$$\sigma^2(g) = \sum_{i=1}^N h_i^2 (\sigma_{ii+}^2)^{-1} + \sum_{j=1}^J h_j^2 (\sigma_{++j}^2)^{-1} + O\left(N_s^{-1} J_s^{-1}\right) \quad \text{as} \quad N \to \infty.$$

Hence, the result of the lemma follows. \hfill \Box

**Lemma 11.** Assume Conditions 1, 2 and 3 hold and $J_s^{-2} N_s (\log N)^2 \to 0$ as $N \to \infty$. If $A_{\theta, \beta} = \{g_i, g_j' : i = 1, ..., N, j = 1, ..., J\}$ such that $g_i, g_j' \in \Omega_N^*$, and $g_i(x) = \theta_i$ and $g_j'(x) = \beta_j$ for $x \in \Omega_N$. Let $C_N = \|A_{\theta, \beta}\|$, the cardinality of $A_{\theta, \beta}$. Then there exist sequences $f_N > 0$ and $d_N \geq 0$ satisfying the followings.

(a). As $N \to \infty$, $f_N^2 / \log C_N \to \infty$.

(b). If $y, v \in \Omega_N$ and $\|y - M_N\|_{\sigma(A_{\theta, \beta})} \leq f_N$, then there exists $n < \infty$ such that for all $N > n$, $\|U_N(y, v)\|_{\sigma(A_{\theta, \beta})} \leq d_N \|y - M_N\|_{\sigma(A_{\theta, \beta})} \|v\|_{\sigma(A_{\theta, \beta})}$. Furthermore, $d_N f_N^2 \to 0$ as $N \to \infty$.

**Proof.** Since $J_s^{-2} N_s (\log N)^2 \to 0$ as $N \to \infty$, there must exists a positive sequence $L_N$ such that
\( L_N \to \infty \) but \( J^{-1} N_*^{1/2} (\log N) L_N \to 0 \) as \( N \to \infty \). Furthermore, note that

\[
\log(C_N) = \log(N + J) \leq \log(2N) = \log(2) + \log(N) = O(\log(N)) \quad \text{as} \quad N \to \infty.
\]

Let \( f_N^2 = \{\log(N)\} L_N \). It is easy to see that the constructed \( f_N \) satisfies part (a) of the lemma.

Now we consider part (b). We seek to find an upper bound for \( \|U_N(y, z)\|_\sigma(A_{\theta, \beta}) \) in order to find \( d_N \) and then show that \( d_N f_N^2 \to 0 \) as \( N \to \infty \). For any \( y, v \in \Omega_N \), by the definition of \( \|\cdot\|_\sigma(A_{\theta, \beta}) \),

\[
\|U_N(y, v)\|_\sigma(A_{\theta, \beta}) = \max_{f \in A_{\theta, \beta}} |f\{U_N(y, v)\}| / \sigma(f).
\]

First note from (68) and (69), we know that for any \( f \in A_{\theta, \beta} \), there exist \( 0 < c_1, c_2 < \infty \) such that for all \( N > n \),

\[
c_1^{-1} N_*^{-\frac{1}{2}} < \sigma(f) < c_2^{-1} J_*^{-\frac{1}{2}}.
\]

So we just need to find an upper bound for \( |f\{U_N(y, v)\}| \) that holds for all \( f \in A_{\theta, \beta} \). Note that

\[
|f\{U_N(y, v)\}| = \sum_{i'=1}^{N} \sum_{j' \in S_j(i')} d_{i'j'}(f) (\sigma^2(y_{i'j'}) - \sigma^2_{i'j'}) |v_{i'j'}| \\
\leq \sum_{i'=1}^{N} \sum_{j' \in S_j(i')} |d_{i'j'}(f)| \cdot |\sigma^2(y_{i'j'}) - \sigma^2_{i'j'}| \cdot |v_{i'j'}|. \tag{72}
\]

Note \( 0 \leq \sigma^2(y_{ij}), \sigma^2_{ij} \leq 1 \), so \( |\sigma^2(y_{ij}) - \sigma^2_{ij}| \leq 1 \). It can be implied that \( |\sigma^2(y_{ij}) - \sigma^2_{ij}| \leq c_3 |f(y - M_N^*)| \) for some positive \( c_3 < \infty \). By the definition of \( \|\cdot\|_\sigma(A_{\theta, \beta}) \), we have \( |f(y - M_N^*)| \leq \|y - M_N^*\|_\sigma(A_{\theta, \beta}) \sigma(f) \). Hence, it follows that for any \( i = 1, \ldots, N, j = 1, \ldots, J, z_{ij} = 1 \),

\[
|\sigma^2(y_{ij}) - \sigma^2_{ij}| \leq c_2^{-1} c_3 J_*^{-\frac{1}{2}} \|y - M_N^*\|_\sigma(A_{\theta, \beta}).
\]

Using a similar strategy, we can show that there exists a positive \( c_4 < \infty \) such that for any \( i = 1, \ldots, N, j = 1, \ldots, J, z_{ij} = 1 \),

\[
|v_{ij}| \leq c_4 J_*^{-\frac{1}{2}} \|v\|_\sigma(A_{\theta, \beta}).
\]
Further, note also that

\[ \sum_{i'=1}^{N} \sum_{j' \in S_j(i')} |d'_{i',j'}(f)| \leq \sum_{i'=1}^{N} \sum_{j' \in S_j(i')} |d'_{i',j'}(f)| + \sum_{i'=1}^{N} \sum_{j' \in S_j(i')} |d''_{i',j'}(f)|. \]

By definition, \( d'_{i',j'} = (\sigma_{i'+}^2)^{-1} w_{i'+} + (\sigma_{j'+}^2)^{-1} w_{j'+} - (\sigma_{++}^2)^{-1} w_{++} \). For any \( f \in A_{\theta, \beta} \), either \( f = g'_j \) or \( f = g_i \). When \( f = g'_j, w_{i'+} = -1/N, \) for \( i' = 1, \ldots, N, w_{j'+} = -1 \) if \( j' = j \) and \( w_{j'+} = 0 \) if \( j' \neq j \), \( w_{++} = -1 \). Hence,

\[ d'_{i',j'}(g'_j) = \begin{cases} -\frac{1}{N}(\sigma_{i'+}^2)^{-1} - (\sigma_{j'+}^2)^{-1} + (\sigma_{++}^2)^{-1} & \text{if } j' = j \\ -\frac{1}{N}(\sigma_{i'+}^2)^{-1} + (\sigma_{++}^2)^{-1} & \text{if } j' \neq j \end{cases} \]

It follows

\[ \sum_{i'=1}^{N} \sum_{j' \in S_j(i')} |d'_{i',j'}(g'_j)| \leq \frac{J^*}{N} \sum_{i'=1}^{N} (\sigma_{i'+}^2)^{-1} + N^*(\sigma_{++}^2)^{-1} + \sum_{i'=1}^{N} \sum_{j' \in S_j(i')} (\sigma_{++}^2)^{-1} \leq c_5, \]

for some positive \( c_5 < \infty \). Furthermore,

\[ \sum_{i'=1}^{N} \sum_{j' \in S_j(i')} |d''_{i',j'}(g'_j)| \leq (N^* J^*)^{\frac{1}{2}} \|d''(g'_j)\|_\sigma \leq c_6, \]

for some positive \( c_6 < \infty \). The last step follows from Condition 1(c) and (68) which implies that \( \|d''(g'_j)\|_\sigma^2 = O \left( N_{\theta, \beta}^{-1} J_{\theta, \beta}^{-1} \right) \). On the other hand, when \( f = g_i \), we have \( w_{i'+} = 1 - 1/N, \) if \( i' = i \), and \( w_{j'+} = -1/N \) for \( i' \neq i \), \( w_{j'+} = 0 \) for all \( j' = 1, \ldots, J \) and \( w_{++} = 0 \). Hence,

\[ d''_{i',j'}(g_i) = \begin{cases} (1 - \frac{1}{N})(\sigma_{i'+}^2)^{-1} & \text{if } i' = i \\ -\frac{1}{N}(\sigma_{i'+}^2)^{-1} & \text{if } i' \neq i \end{cases} \]

It follows

\[ \sum_{i'=1}^{N} \sum_{j' \in S_j(i')} |d''_{i',j'}(g_i)| = \sum_{j' \in S_j(i')} \left( 1 - \frac{1}{N} \right)(\sigma_{i'+}^2)^{-1} - \sum_{j' \in S_j(i')} \frac{1}{N}(\sigma_{i'+}^2)^{-1} \leq c_7, \]

71
for some positive $c_7 < \infty$. Furthermore,

$$\sum_{i'=1}^{N} \sum_{j' \in S_j(i')} |d''_{i',j'}(g_i)| \leq (N^* J^*)^{1/2} \|d''(g_i)\|_{\sigma} \leq c_8,$$

for some positive $c_8 < \infty$. The last step follows from Condition 1(c) and (69) which implies that $\|d''(g_i)\|_{\sigma}^2 = O \left(N_s^{-1} J_s^{-1}\right)$. Overall,

$$\|U_N(y, v)\|_{\sigma}(A_{\theta, \beta}) = \max_{f \in A_{\theta, \beta}} |f \{U_N(y, v)\}| \sigma(f) \leq \max_{f \in A_{\theta, \beta}} |f \{U_N(y, v)\}| \max_{f \in A_{\theta, \beta}} \{\sigma(f)^{-1}\} \leq c_1 c_2^{-2} c_3 c_4 \max\{c_5 + c_6, c_7 + c_8\} J_*^{-1} N_*^{1/2} \|y - M_N\|_{\sigma}(A_{\theta, \beta}) \|v\|_{\sigma}(A_{\theta, \beta}).$$

Note that in this case we can take $d_N = c_1 c_2^{-2} c_3 c_4 \max\{c_5 + c_6, c_7 + c_8\} J_*^{-1} N_*^{1/2}$. We have

$$d_N f_N^2 = c_1 c_2^{-2} c_3 c_4 \max\{c_5 + c_6, c_7 + c_8\} J_*^{-1} N_*^{1/2} \log(N) L_N \to 0 \quad \text{as} \quad N \to \infty.$$

Hence both parts (a) and (b) of the lemma are satisfied.

A.3 Senator Rankings

With the set-up in Section 5.2 in the main article, we present a full list of rankings for senators serving the 111th, the 112th and the 113th United States senate according to their conservativeness scores. The results are summarized in Tables 3 and 4 below. We observe from Table 3 that all the top 62 most conservative senators predicted by the model are Republicans. While the Democrats and the independent politicians are predicted to have much lower conservativeness scores as presented in Table 4. This aligns well with the public perceptions about the Republican party and the Democratic party. Standard errors of the estimated row parameters (i.e. senator’s conservativeness score) are also included to facilitate inferences.
Table 3: Ranking of the top 62 most conservative senators predicted by the model. Rep represents
the Republican party and states are listed in their standard abbreviations. \( \hat{\theta} \) represents the con-
servativeness score of senators and \( \text{s.e.}(\hat{\theta}) \) is the standard error of the estimated conservativeness score.

| Rank | Senator | State | Party | \( \hat{\theta} \) | s.e.(\( \hat{\theta} \)) | Rank | Senator | State | Party | \( \hat{\theta} \) | s.e.(\( \hat{\theta} \)) |
|------|---------|-------|-------|-------------|----------------|------|---------|-------|-------|-------------|----------------|
| 1    | Demint  | SC    | Rep   | 5.87       | 0.157         | 2    | Lee     | UT    | Rep   | 5.73       | 0.138         |
| 3    | Cruz    | TX    | Rep   | 5.65       | 0.195         | 4    | Coburn  | OK    | Rep   | 5.25       | 0.114         |
| 5    | Paul    | KY    | Rep   | 5.24       | 0.129         | 6    | Scott   | SC    | Rep   | 5.17       | 0.176         |
| 7    | Bunning | KY    | Rep   | 4.92       | 0.204         | 8    | Johnson | WI    | Rep   | 4.84       | 0.119         |
| 9    | Risch   | ID    | Rep   | 4.81       | 0.102         | 10   | Inhofe  | OK    | Rep   | 4.69       | 0.103         |
| 11   | Crapo   | ID    | Rep   | 4.56       | 0.097         | 12   | Sessions | AL    | Rep   | 4.48       | 0.096         |
| 13   | Enzi    | WY    | Rep   | 4.36       | 0.094         | 14   | Barasso | WY    | Rep   | 4.35       | 0.094         |
| 15   | Cornyn  | TX    | Rep   | 4.33       | 0.095         | 16   | Rubio   | FL    | Rep   | 4.25       | 0.112         |
| 17   | Ensign  | NV    | Rep   | 4.24       | 0.166         | 18   | Vitter  | LA    | Rep   | 4.20       | 0.094         |
| 19   | Fischer | NE    | Rep   | 4.14       | 0.145         | 20   | Toomey  | PA    | Rep   | 4.12       | 0.109         |
| 21   | Kyl     | AZ    | Rep   | 4.10       | 0.115         | 22   | Roberts | KS    | Rep   | 4.06       | 0.091         |
| 23   | Mcconnell| KY    | Rep   | 4.02       | 0.089         | 24   | Thune   | SD    | Rep   | 3.95       | 0.088         |
| 25   | Burr    | NC    | Rep   | 3.95       | 0.090         | 26   | Moran   | KS    | Rep   | 3.89       | 0.109         |
| 27   | Grassley| IA    | Rep   | 3.80       | 0.086         | 28   | Shelby  | AL    | Rep   | 3.78       | 0.086         |
| 29   | Boozman | AR    | Rep   | 3.68       | 0.105         | 30   | Chambliss| GA   | Rep   | 3.65       | 0.087         |
| 31   | Mc Cain | AZ    | Rep   | 3.65       | 0.086         | 32   | Brownback| KS  | Rep   | 3.61       | 0.153         |
| 33   | Coats   | IN    | Rep   | 3.51       | 0.101         | 34   | Johans  | NE    | Rep   | 3.39       | 0.082         |
| 35   | Isakson | GA    | Rep   | 3.38       | 0.082         | 36   | Hatch   | UT    | Rep   | 3.38       | 0.083         |
| 37   | Lemieux | FL    | Rep   | 3.34       | 0.188         | 38   | Blunt   | MO    | Rep   | 3.31       | 0.099         |
| 39   | Wicker  | MS    | Rep   | 3.29       | 0.080         | 40   | Portman | OH    | Rep   | 3.28       | 0.098         |
| 41   | Corker  | TN    | Rep   | 3.27       | 0.080         | 42   | Heller  | NV    | Rep   | 3.26       | 0.100         |
| 43   | Hutchison| TX    | Rep   | 3.25       | 0.105         | 44   | Graham  | SC    | Rep   | 3.18       | 0.080         |
| 45   | Flake   | AZ    | Rep   | 3.03       | 0.125         | 46   | Ayotte  | NH    | Rep   | 3.02       | 0.095         |
| 47   | Hoeven  | ND    | Rep   | 2.97       | 0.094         | 48   | Bennett | UT    | Rep   | 2.74       | 0.127         |
| 49   | Alexander| TN   | Rep   | 2.71       | 0.075         | 50   | Kirk    | IL    | Rep   | 2.67       | 0.105         |
| 51   | Cochran | MS    | Rep   | 2.63       | 0.075         | 52   | Chiesa  | NJ    | Rep   | 2.61       | 0.343         |
| 53   | Gregg   | NH    | Rep   | 2.59       | 0.127         | 54   | Martinez| FL    | Rep   | 2.47       | 0.186         |
| 55   | Lugar   | IN    | Rep   | 2.29       | 0.088         | 56   | Bond    | MO    | Rep   | 2.25       | 0.118         |
| 57   | Murkowski| AK   | Rep   | 1.47       | 0.066         | 58   | Brown   | MA    | Rep   | 1.29       | 0.103         |
| 59   | Voinovich| OH   | Rep   | 1.22       | 0.102         | 60   | Snowe   | ME    | Rep   | 1.06       | 0.080         |
| 61   | Specter | PA    | Rep   | 1.03       | 0.192         | 62   | Collins | ME    | Rep   | 0.82       | 0.064         |
Table 4: Ranking of the top 63-139 most conservative senators predicted by the model. Dem and Ind represent the Democratic party and independent politician, respectively. States are presented in their standard abbreviations. $\hat{\theta}$ represents the conservativeness score of senators and $\text{s.e.}(\hat{\theta})$ is the standard error of the estimated conservativeness score.

| Rank | Senator  | State | Party | $\theta$ | s.e.$(\hat{\theta})$ | Rank | Senator  | State | Party | $\theta$ | s.e.$(\hat{\theta})$ |
|------|----------|-------|-------|---------|-----------------|------|----------|-------|-------|---------|-----------------|
| 63   | Nelson   | NE    | Dem   | -0.05   | 0.084           | 64   | Bayh     | IN    | Dem   | -0.13   | 0.104           |
| 65   | Manchin  | WV    | Dem   | -0.66   | 0.099           | 66   | Feingold | WI    | Dem   | -0.92   | 0.115           |
| 67   | Lincoln  | AR    | Dem   | -0.96   | 0.119           | 68   | Mccaskill| MO    | Dem   | -1.15   | 0.083           |
| 69   | Webb     | VA    | Dem   | -1.49   | 0.108           | 70   | Pryor    | AR    | Dem   | -1.63   | 0.094           |
| 71   | Lieberman| CT    | Dem   | -1.68   | 0.113           | 72   | Heitkamp | ND    | Dem   | -1.87   | 0.183           |
| 73   | Donnelly | IN    | Dem   | -1.87   | 0.182           | 74   | Hagan    | NC    | Dem   | -1.90   | 0.100           |
| 75   | Byrd     | WV    | Dem   | -2.00   | 0.217           | 76   | Warner   | VA    | Dem   | -2.06   | 0.105           |
| 77   | Landrieu | LA    | Dem   | -2.07   | 0.106           | 78   | Tester   | MT    | Dem   | -2.11   | 0.105           |
| 79   | Baucus   | MT    | Dem   | -2.11   | 0.112           | 80   | Bennet   | CO    | Dem   | -2.16   | 0.107           |
| 81   | Klobuchar| MN    | Dem   | -2.26   | 0.109           | 82   | Conrad   | ND    | Dem   | -2.29   | 0.131           |
| 83   | King     | ME    | Ind   | -2.30   | 0.208           | 84   | Nelson   | FL    | Dem   | -2.32   | 0.112           |
| 85   | Kohl     | WI    | Dem   | -2.34   | 0.131           | 86   | Carper   | DE    | Dem   | -2.36   | 0.112           |
| 87   | Udall    | CO    | Dem   | -2.39   | 0.113           | 88   | Begich   | AK    | Dem   | -2.43   | 0.116           |
| 89   | Dorgan   | ND    | Dem   | -2.44   | 0.167           | 90   | Reid     | NV    | Dem   | -2.68   | 0.122           |
| 91   | Shaheen  | NH    | Dem   | -2.76   | 0.125           | 92   | Kaine    | VA    | Dem   | -2.80   | 0.246           |
| 93   | Casey    | PA    | Dem   | -2.83   | 0.127           | 94   | Cantwell | WA    | Dem   | -2.84   | 0.127           |
| 95   | Coons    | DE    | Dem   | -2.84   | 0.170           | 96   | Specter  | PA    | Dem   | -2.84   | 0.222           |
| 97   | Walsh    | MT    | Dem   | -2.85   | 0.395           | 98   | Wyden    | OR    | Dem   | -2.97   | 0.132           |
| 99   | Bingaman | NM    | Dem   | -3.03   | 0.155           | 100  | Johnson  | SD    | Dem   | -3.09   | 0.137           |
| 101  | Stabenow | MI    | Dem   | -3.11   | 0.137           | 102  | Cowan    | MA    | Dem   | -3.19   | 0.439           |
| 103  | Merkley  | OR    | Dem   | -3.19   | 0.140           | 104  | Sanders  | VT    | Ind   | -3.23   | 0.143           |
| 105  | Feinstein| CA    | Dem   | -3.24   | 0.143           | 106  | Perry    | MA    | Dem   | -3.25   | 0.165           |
| 107  | Kaufman  | DE    | Dem   | -3.28   | 0.219           | 108  | Murray   | WA    | Dem   | -3.29   | 0.143           |
| 109  | Heinrich | NM    | Dem   | -3.30   | 0.290           | 110  | Menendez | NJ    | Dem   | -3.32   | 0.144           |
| 111  | Inouye   | HI    | Dem   | -3.33   | 0.169           | 112  | Boxer    | CA    | Dem   | -3.35   | 0.148           |
| 113  | Dodd     | CT    | Dem   | -3.38   | 0.218           | 114  | Warren   | MA    | Dem   | -3.45   | 0.307           |
| 115  | Levin    | MI    | Dem   | -3.52   | 0.152           | 116  | Blumenthal| CT   | Dem   | -3.52   | 0.214           |
| 117  | Kirk     | MA    | Dem   | -3.54   | 0.716           | 118  | Akaka    | HI    | Dem   | -3.54   | 0.174           |
| 119  | Franken  | MN    | Dem   | -3.55   | 0.166           | 120  | Rockefeller| WV  | Dem   | -3.56   | 0.161           |
| 121  | Mikulski | MD    | Dem   | -3.60   | 0.158           | 122  | Leahy    | VT    | Dem   | -3.63   | 0.158           |
| 123  | Harkin   | IA    | Dem   | -3.64   | 0.158           | 124  | Lautenberg| NJ   | Dem   | -3.65   | 0.179           |
| 125  | Schumer  | NY    | Dem   | -3.65   | 0.159           | 126  | Reed     | RI    | Dem   | -3.67   | 0.157           |
| 127  | Gillibrand| NY    | Dem   | -3.67   | 0.158           | 128  | Murphy   | CT    | Dem   | -3.68   | 0.327           |
| 129  | Markey   | MA    | Dem   | -3.73   | 0.465           | 130  | Whitehouse| RI   | Dem   | -3.74   | 0.163           |
| 131  | Cardin   | MD    | Dem   | -3.82   | 0.163           | 132  | Durbin   | IL    | Dem   | -3.83   | 0.164           |
| 133  | Udall    | NM    | Dem   | -3.85   | 0.165           | 134  | Brown    | OH    | Dem   | -3.89   | 0.168           |
| 135  | Baldwin  | WI    | Dem   | -3.90   | 0.352           | 136  | Booker   | NJ    | Dem   | -4.14   | 0.572           |
| 137  | Hirono   | HI    | Dem   | -4.17   | 0.383           | 138  | Burris   | IL    | Dem   | -4.43   | 0.297           |
| 139  | Schatz   | HI    | Dem   | -4.74   | 0.468           |