Specification and Estimation of Matrix Exponential Social Network Models with an Application to Add Health Data*

Ye Yang† Osman Doğan‡ Süleyman Taşpınar§

April 6, 2025

Abstract

In this paper, we introduce a new specification for modeling peer effects due to social interactions. We refer to this specification as the matrix exponential social (MES) network model. The MES network model allows for the endogenous effect, the contextual effects, heterogeneity across groups, the correlation in unobserved characteristics of members, as well as an unknown form of heteroskedasticity. We propose consistent estimation and inference methods for the MES network model. In an extensive simulation study, we show that the proposed methods perform satisfactorily. In an empirical application using the Add Health data, we illustrate how the MES network model can be used in identifying peer effects on academic success, recreational activities and smoking frequency of adolescents.

JEL-Classification: C13, C21, I21.
Keywords: Social interaction, social network models, heteroskedasticity.

*This research uses data from Add Health, funded by grant P01 HD31921 (Harris) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), with cooperative funding from 23 other federal agencies and foundations. Add Health is currently directed by Robert A. Hummer and funded by the National Institute on Aging cooperative agreements U01 AG071448 (Hummer) and U01AG071450 (Aiello and Hummer) at the University of North Carolina at Chapel Hill. Add Health was designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill.

†School of Accounting, Capital University of Economics and Business, Beijing, China, email: yeyang557@hotmail.com

‡Department of Economics, Istanbul Technical University, Istanbul, Türkiye, email: osmandogan@itu.edu.tr

§Department of Economics, Queens College, CUNY, New York, U.S.A., email: staspinar@qc.cuny.edu

1 Introduction

Peer effects refer to the influence on an individual's behavior or decision resulting from the individual's interactions with his/her peers. There is a growing body of empirical literature showing that peer effects exist in a wide range of settings, including education, crime, obesity, smoking, sports, tax compliance and evasion, and program participation. Various social interaction models proposed for estimating peer effects allow for a group member's outcome to depend on both the outcomes and characteristics of other group members. One main difficulty for the empirical analysis based on these models is that it may not be possible to distinguish the correlation of outcomes due to social interactions from the correlation due to the unobserved group characteristics. Another major obstacle is to distinguish the impact of peers' outcomes from the impact of peers' characteristics, which is known as the reflection problem (Blume et al., 2015; Manski, 1993; Moffitt, 2001).

Similarities of these models to spatial autoregressive models from spatial econometrics have been recognized by several studies (Bramoullé et al., 2009, 2020; Lee, 2007; Lee et al., 2010; Lin, 2010, 2015; Liu and Lee, 2010). This strand of the literature shows that the spatial autoregressive network models may resolve the reflection problem and allow for the identification of endogenous effect, contextual effects, and unobserved correlation effect. In these types of models, the endogenous peer effect and the correlation of unobserved characteristics of members are specified through spatial autoregressive processes. Following this strand of the literature, we introduce a new specification that allows a group member's outcome to depend on (i) the outcomes of the other members of the group through a matrix exponential term, (ii) the characteristics of the other members through a spatial lag term, (iii) the unobserved group heterogeneity, and (iv) the correlation in the unobserved characteristics of members specified through a matrix exponential term. We will refer to this model as the matrix exponential social (MES) network model.

We show that the best reply function based on two alternative quadratic utility functions that account for all externality levels generated by group members heterogeneity or all spillover effects due to social interactions within the group can yield the MES network model. The MES network model has two important features. First, the endogenous peer effects and the correlation in the unobserved group characteristics are specified through matrix exponential terms rather than spatial autoregressive processes. It is well known in the literature that the models specified in terms of matrix exponential terms enjoy several properties in terms of specification and estimation issues (Chiu et al., 1996; Debarsy et al., 2015; LeSage and Pace, 2007; LeSage and Pace, 2009; Yang et al., 2021, 2022, 2024). Because the matrix exponential terms are always invertible, our specification has a reduced (or equilibrium) form, and does not require any restrictions on the endogenous and correlated effects. On the other hand, the estimation of the spatial autoregressive network models requires a restrictive parameter space for the endogenous and correlated effects. Moreover, the likelihood based estimation of our network model has the computational advantage because the likelihood function does not involve any matrix determinant terms that need to be computed in each iteration during the estimation.

The second feature of our specification is that it allows for potential heteroskedasticity in the unobserved idiosyncratic characteristics (i.e., the error terms). This means that the outcomes of group members are allowed to be affected by the unobserved factors in different ways. Given that group members can differ in many ways, it is natural to expect that the homoskedasticity assumption generally will not hold in empirical applications. However, the literature has only considered spatial autoregressive network models under the homoskedasticity assumption (Bramoullé et al., 2020; Lee, 2007; Lee et al., 2010; Lin, 2010, 2015; Liu and Lee, 2010). Importantly, the estimators suggested in these studies may not be consistent under heteroskedasticity because these estimators are usually derived from some non-linear objective functions. As in White (1980), our approach does not require a specific formal model for the structure of heteroskedasticity and allows for heteroskedasticity-robust inference.

In this paper, we explore the likelihood based estimation of the MES network model under both homoskedastic and heteroskedastic error term cases. We consider two estimation approaches: (i) a transformation approach and (ii) a direct approach. In the transformation approach, we first eliminate the group fixed effects from the model using a orthogonal transformation based on the decomposition of a projection matrix (Lee and Yu, 2010). In the homoskedastic case, we show that the quasi maximum likelihood estimator (QMLE) defined based on the likelihood function of the transformed model attains the usual large sample properties under some regularity conditions. In the heteroskedastic case, we show that the QMLE may not be consistent as the expectation of the score functions evaluated at the true parameter vector may not be zero. Therefore, our estimation strategy dwells on re-centering the score functions analytically to ensure that their expectations become zero. We then defined a robust M-estimator (RME) as the root of the adjusted score functions and establish its large sample properties under some assumptions.

The transformation approach is only applicable for models that have row normalized network matrices. Therefore, we also consider a direct estimation approach that does not necessitate the row-normalization. In the direct approach, we estimate the group fixed effects along with the main parameter vector based on the likelihood function of the original model. In both homoskedastic and heteroskedastic cases, we show how to adjust score functions such that their expectations evaluated at the true parameter vector are zero in all cases. Using suitably adjusted score functions, we propose an M-estimator (ME) in the homoskedastic case and an RME in the heteroskedastic case. We formally establish consistency and asymptotic normality of both estimators under some assumptions.

The variance-covariance matrix of all suggested estimators takes a sandwich form. For the QMLE based on the transformation approach, we suggest using a plug-in estimator and show how consistent plug-in estimators can be formulated for the third and fourth moments of the error term. In the case of RME under the transformation approach, and ME and RME under the direct approach, we show that the expectation of the Hessian matrix can be consistently estimated by its sample counterpart evaluated at a consistent estimator. For the variance-covariance matrix of the adjusted score functions, we show that a plug-in estimator is consistent in the case of the RME based on the transformation approach. However, in the case of both ME and RME based on the direct approach, the plug-in estimators of the variance-covariance matrix of the adjusted score functions have asymptotic bias because of the estimation of the group fixed effects. We suggest analytical bias corrections for both estimators.

In an extensive Monte Carlo study, we assess the finite sample properties of the proposed estimators and inference methods under both homoskedastic and heteroskedastic settings. Our results show that the QMLE based on the transformation approach performs satisfactorily under homoskedasticity. However, under heteroskedasticity, while the QMLE shows satisfactory performance in terms of bias, it can report empirical coverage rates that are lower than the nominal value. Our results show that the ME based on the direct approach performs similar to the QMLE. In particular, it exhibits negligible bias but can report empirical coverage rates that are lower than the nominal value for some parameters under heteroskedasticity. Our simulation results also show that the RME based on both transformation and direct approaches perform satisfactorily under both homoskedastic and heteroskedastic settings.

In an empirical application, we estimate the MES network model using data from the National Longitudinal Study of Adolescent to Adult Health (Add Health) to investigate peer effects on academic achievement, participation in recreational activities, and smoking. The Add Health data sets have been extensively used to explore the role of peer effects on academic achievement, study effort, school activities, delinquency, and risky behaviors. See, among others, (Boucher et al., 2024; Calvó-Armengol et al., 2009; Clark and Lohéac, 2007; Fruehwirth, 2014; Goldsmith-Pinkham and Imbens, 2013; Hsieh and Lee, 2016; Hsieh and Lin, 2021; Lee et al., 2014; Lee et al., 2021; Lin, 2010, 2015; Zenou and Patacchini, 2012). For academic achievement and participation in recreational activities, our results provide statistical evidence only for the correlation in unobserved factors (the correlation effect), while for smoking frequency, our results show statistical evidence for both endogenous and correlation effects.

Methodologically, our utility functions that yield the MES network model are related to the conformist and spillover utility functions that underpin the linear-in-means model of Manski (1993) (Boucher et al., 2024; Brock and Durlauf, 2001; Patacchini and Zenou, 2009). In the spillover model, an individual always benefits from the average peers' efforts (peers' social norm), while in the conformist model, the individual incurs a cost for deviating from the social norm. Boucher et al. (2024) recently suggest a general utility function that yields a best reply function that is non-linear in the social norm and encompasses both the spillover and conformist models as special cases. We suggest two alternative utility functions, where the first one accounts for all externality levels generated by group members' heterogeneity and the second one accounts for all spillover effects due to social interactions within the group.

Our paper is also closely related to the research on estimation and inference methods for spatial autoregressive network models. Lee (2007) is the first to consider a spatial autoregressive model in a group setting, allowing for endogenous group interactions, contextual factors, and group-specific fixed effects (i.e., unobserved group heterogeneity). Using the Add Health data, Lin (2010, 2015) employ the spatial autoregressive network model with a network matrix based on friendship links to estimate peer effects on adolescent developmental outcomes. Subsequently, Bramoullé et al. (2009) investigate the identification of network models and demonstrate that identification depends on the structure of the network matrix. Lee et al. (2010) and Liu and Lee (2010) extend these models with general network matrices in the quasi-maximum likelihood and generalized method of moments settings, respectively. See Bramoullé et al. (2020) for an extensive review of peer effects in networks.

The rest of the paper is organized as follows. In Section 2, we briefly describe a setting in which the matrix exponential social network model can arise. Section 3 presents the details of the transformation approach. We establish the large sample properties of the QMLE under homoskedasticity in Section 3.1 and of the RME under heteroskedasticity in Section 3.2. In Section 4, we consider the direct approach and define the ME and RME for the homoskedastic and heteroskedastic cases, respectively. In Section 5, we investigate the finite sample performance of the proposed estimators through extensive simulations. In Section 6, we provide our empirical application using the Add Health data. Finally, Section 7 offers concluding remarks. All technical details are provided in a supplementary web appendix.

2 Model specification

In this section, we show how our model specification can arise from a utility maximization framework. We consider n agents partitioned into R networks such that there are n_r agents in the rth network for $r = 1, 2, \ldots, R$. Let $W_r = (w_{ij})$ be the $n_r \times n_r$ network matrix for the rth group such that $w_{ij,r} = 1$ if i and j are connected, and zero otherwise. The degree of the ith agent is given by $\sum_{j=1}^{n_r} w_{ij,r}$, and we do not require that each agent has the same degree. In particular, W_r may or may not be directed or row-normalized. Let y_{ir} be the effort level (for an outcome variable) of the ith agent in group r, and π_{ir} denote the (ex ante) heterogeneity of the ith agent. Let $Y_r = (y_{1r}, y_{2r}, \ldots, y_{n_r r})'$ and $\Pi_r = (\pi_{1r}, \pi_{2r}, \ldots, \pi_{n_r r})'$. For the given network structure W_r and heterogeneity Π_r, we assume that the ith agent chooses its effort level to maximize the following utility function:

$$\mathcal{U}_{ir}(y_{ir}, W_r, \Pi_r) = y_{ir} e_{ir}^{'} \left(\sum_{j=0}^{\infty} \frac{\rho_0^j}{j!} W_r^j \Pi_r \right) - \frac{1}{2} y_{ir}^2, \tag{2.1}$$

where e_{ir} is the ith column of the $n_r \times n_r$ identify matrix I_{n_r} and ρ_0 is scalar parameter. In $\mathcal{U}_{ir}(y_{ir}, W_r, \Pi_r)$, the first term represents the benefit, while the second quadratic term represents the cost of choosing the effort level y_{ir}. The benefit term combines the agent's own effort level y_{ir} with the ith element of the infinite sum $\sum_{j=0}^{\infty} \frac{\rho_0^j}{j!} W_r^j \Pi_r$, which is defined as $\sum_{j=0}^{\infty} \frac{\rho_0^j}{j!} W_r^j \Pi_r = \left(\Pi_r + \rho W_r \Pi_r + \frac{\rho_0^2}{2} W_r^2 \Pi_r + \frac{\rho_0^3}{6} W_r^3 \Pi_r + \ldots \right)$. This sum reflects the externalities generated by heterogeneity due to social interactions within the rth group. The first term in the sum, Π_r, represents the externality effect of the agent's own heterogeneity; the second term, $\rho_0 W_r \Pi_r$, represents the externality level from first-order neighbors; the third term, $\frac{\rho_0^2}{2} W_r^2 \Pi_r$, represents the externality level from second-order neighbors; and all remaining terms capture the externalities from higher-order neighbors' heterogeneity.

Note that our utility function in (2.1) does not depend on the effort levels of other agents. To allow for spillover effects, we can alternatively assume the following utility function:

$$\mathcal{U}_{ir}(y_{ir}, Y_{-i,r}, W_r, \Pi_r) = y_{ir} e_{ir}^{'} \left(\Pi_r - \frac{1}{2} \sum_{j=1}^{\infty} \frac{\rho_1^j}{j!} W_r^j Y_r \right) - \frac{1}{2} y_{ir}^2, \tag{2.2}$$

where $Y_{-i,r} = (y_{1r}, y_{2r}, \ldots, y_{i-1,r}, y_{i+1,r}, \ldots, y_{n_r r})'$ is the vector of effort levels, excluding the effort level of the ith agent and ρ_1 is a scalar parameter. In this formulation, the infinite sum $\sum_{j=1}^{\infty} \frac{\rho_1^j}{j!} W_r^j Y_r = \left(\rho_1 W_r Y_r + \frac{\rho_1^2}{2} W_r^2 Y_r + \frac{\rho_1^3}{6} W_r^3 Y_r + \ldots \right)$ represents all spillover effects generated by social interactions within the rth group. The first term in the sum is the social norm that represents the spillover effect from the first-order neighbors, the second term represents the spillover effect from the second-order neighbors, and all other terms represent the spillover effects from high-order neighbors. Note that if $\rho_0 = \rho_1 = 0$, then $\mathcal{U}_{ir}(y_{ir}, Y_{-i,r}, W_r, \Pi_r) = \mathcal{U}_{ir}(y_{ir}, W_r, \Pi_r)$, and social interactions within the rth group do not generate externalities or spillover effects beyond the agent's own heterogeneity. In either case, the benefit from the externality and spillover effects depends on the sign and magnitude of the scalar parameters ρ_0 and ρ_1.

Our utility functions are different from the spillover and conformist utility functions that are used to derive the linear-in-means model of Manski (1993). In the spillover model, the ith agent chooses its effort level by maximizing the following utility function:

$$\mathcal{U}^S_{ir}(y_{ir}, Y_{-i,r}, W_r, \Pi_r) = y_{ir}\pi_{ir} + \rho^S y_{ir} e^{'}_{ir} W_r Y_r - \frac{1}{2} y^2_{ir}, \tag{2.3}$$

where $0 \le \rho^S < 1$ is a parameter measuring the intensity of the spillover effect of the social norm $e^{'}_{ir} W_r Y_r$ (Boucher et al., 2024; Brock and Durlauf, 2001). In the conformist model, the ith agent chooses its effort level by maximizing the following utility function:

$$\mathcal{U}^C_{ir}(y_{ir}, Y_{-i,r}, W_r, \Pi_r) = y_{ir}\pi_{ir} - \frac{\rho^C}{2}\left(y_{ir} - e^{'}_{ir} W_r Y_r\right)^2 - \frac{1}{2} y^2_{ir}, \tag{2.4}$$

where $\rho^C \ge 0$ is a scalar parameter for taste conformity(Boucher et al., 2024; Patacchini and Zenou, 2009). Recently, Boucher et al. (2024) generalize both models and assume that the ith agent chooses its effort level by maximizing the following utility function:

$$\mathcal{U}^G_{ir}(y_{ir}, Y_{-i,r}, W_r, \Pi_r) = y_{ir}\pi_{ir} + \rho^G_1 y_{ir} \tilde{y}_{-i,r}(\beta) - \frac{1}{2}\left(y^2_{ir} + \rho^G_2 \left(y_{ir} - \tilde{y}_{-i,r}(\beta)\right)^2\right), \tag{2.5}$$

where $\tilde{y}_{-i,r}(\beta) = \left(\sum_{j=1}^{n} w_{ij,r} y^{\beta}_{jr}\right)^{1/\beta}$ is the CES social norm with the unrestricted parameter β, and ρ^G_1 and ρ^G_2 are scalar parameters. In $\mathcal{U}^G_{ir}(y_{ir}, Y_{-i,r}, W_r, \Pi_r)$, the sum of the first two terms denote the benefit, while the last term the cost of choosing y_{ir}. Under certain constraints on ρ^G_1 and ρ^G_2, Boucher et al. (2024) show that $\mathcal{U}^G_{ir}(y_{ir}, Y_{-i,r}, W_r, \Pi_r)$ yields a best response function for y_{ir}.

Our approach differs from these approaches in two key ways. First, in $\mathcal{U}_{ir}(y_{ir}, W_r, \Pi_r)$, we account for all externality levels generated by heterogeneity, and in $\mathcal{U}_{ir}(y_{ir}, Y_{-i,r}, W_r, \Pi_r)$, we account for all spillover effects due to social interactions within the rth group. Second, we do not impose any restrictions on the scalar parameters ρ_0 and ρ_1, or the network matrix W_r. In our setting, the best reply function based on either $\mathcal{U}_{ir}(y_{ir}, W_r, \Pi_r)$ or $\mathcal{U}_{ir}(y_{ir}, Y_{-i,r}, W_r, \Pi_r)$

is derived as

$$y_{ir} = e_{ir}^{'} e^{\varphi W_r} \Pi_r, \tag{2.6}$$

where $e^{\varphi W_r} = \sum_{j=0}^{\infty} \frac{\varphi^j}{j!} W_r^j$ is the matrix exponential term, and $\varphi = \rho_0$ in the case of $\mathcal{U}_{ir}(y_{ir}, W_r, \Pi_r)$, while $\varphi = -\rho_1$ in the case of $\mathcal{U}_{ir}(y_{ir}, Y_{-i,r}, W_r, \Pi_r)$. Therefore, in our case, the best reply function for the rth group is simply given by

$$Y_r = e^{\varphi W_r} \Pi_r. \tag{2.7}$$

We further assume that the individual heterogeneity across the members of group r is given by

$$\Pi_r = X_r \beta_1 + W_r X_r \beta_2 + \lambda_r \mathbf{1}_{n_r} + U_r, \tag{2.8}$$

where X_r is the $n_r \times k_x$ matrix of exogenous observed own characteristics with the matching parameter vector β_1, $W_r X_r$ is the $n_r \times k_x$ matrix of contextual variables (peers' observed characteristics) with the matching parameter vector β_2, λ_r is the group-specific fixed effect for the rth group, $\mathbf{1}_{n_r}$ is the $n_r \times 1$ vector of ones, and $U_r = (u_{1r}, \ldots, u_{n_r r})'$ is the $n_r \times 1$ vector of unobserved characteristics. We further allow for the possibility that the unobserved characteristics are correlated among the group members by assuming that $e^{\tau M_r} U_r = V_r$, where τ is a scalar parameter, M_r is another $n_r \times n_r$ network matrix specifying links among the unobserved factors, and $V_r = (v_{1r}, \ldots, v_{n_r r})'$ is the $n_r \times 1$ vector of unobserved idiosyncratic characteristics (i.e., the error terms).

Hence, our suggested MES network model takes the following form:

$$e^{\alpha W_r} Y_r = X_r \beta_1 + W_r X_r \beta_2 + \lambda_r \mathbf{1}_{n_r} + U_r, \quad e^{\tau M_r} U_r = V_r, \tag{2.9}$$

where $\alpha = -\varphi$. In (2.9), α is the endogenous effect, β_1 is the vector of own effects from own observed characteristics, β_2 is the vector of contextual effects from the observed peers' characteristics, λ_r is the unobserved group heterogeneity, and τ is the correlation effect. Since the matrix exponential terms $e^{\alpha W_r}$ and $e^{\tau M_r}$ are always invertible for any finite value of α and τ, the reduced form of our model always exists and is given by

$$Y_r = e^{-\alpha W_r} \left(X_r \beta_1 + W_r X_r \beta_2 + \lambda_r \mathbf{1}_{n_r} + e^{-\tau M_r} V_r \right). \tag{2.10}$$

3 Transformation approach

3.1 Estimation under homoskedasticity

In this section, we provide the details of our estimation approach for the MES network model under homoskedasticity. We require the following assumptions.

Assumption 1. *The v_{ir}'s are independent and identically distributed (i.i.d.) across i and r with mean zero and variance σ_0^2, and $\mathrm{E}|v_{ir}|^{4+\varrho} < \infty$ for some $\varrho > 0$.*

Assumption 2. *The network matrices are row-normalized, i.e., $W_r \mathbf{1}_{n_r} = \mathbf{1}_{n_r}$ and $M_r \mathbf{1}_{n_r} = \mathbf{1}_{n_r}$ for $r = 1, \dots, R$.*

Assumption 3. *The network matrices W_r and M_r are uniformly bounded in both row sum and column sum matrix norms for $r = 1, \dots, R$.*

Assumption 4. *There exists a constant $c > 0$ such that $|\alpha| \leq c$ and $|\tau| \leq c$, and the true parameter vector $\zeta_0 = (\alpha_0, \tau_0)'$ lies in the interior of $\Delta = [-c, c] \times [-c, c]$.*

Under Assumption 1, we assume that the error terms are independent and identically distributed across i and r. The moment condition $\mathrm{E}|v_{ir}|^{4+\varrho} < \infty$ ensures that we can apply the central limit theorem (CLT) for linear and quadratic forms to the score functions of our model (Kelejian and Prucha, 2001, 2010). Assumption 2 is necessary when introducing the likelihood function. Assumption 3 provides the essential properties of the network matrices. It ensures that the network correlation is limited to a manageable degree (Kelejian and Prucha, 2001, 2010). Assumption 4 requires that the parameter space of the parameters in the matrix exponential terms is compact. This assumption and Assumption 3 imply that the matrix exponential terms are uniformly bounded in both row sum and column sum matrix norms. This can be seen from $\left\| e^{\alpha W_r} \right\| = \| \sum_{i=0}^{\infty} \alpha^i W_r^i / i! \| \leq \sum_{i=0}^{\infty} |\alpha|^i \|W_r\|^i / i! = e^{|\alpha| \|W_r\|}$, which is bounded if $|\alpha|$ and $\|W_r\|$ are bounded, where $\| \cdot \|$ is either the row sum or the column sum matrix norm.

Define the $n_r \times k$ matrix $Z_r = (X_r, W_r X_r)$ and the $k \times 1$ vector $\beta = (\beta_1', \beta_2')'$, where $k = 2k_x$. Then, the MES network model can be written as

$$e^{\alpha W_r} Y_r = Z_r \beta + \lambda_r \mathbf{1}_{n_r} + U_r, \quad e^{\tau M_r} U_r = V_r, \quad r = 1, 2, \dots, R. \tag{3.1}$$

In order to avoid the incidental parameter problem in the estimation of (3.1), we need to eliminate the group fixed effects from the model. To that end, we use an orthogonal transformation based on the decomposition of $J_r = I_{n_r} - \frac{1}{n_r} \mathbf{1}_{n_r} \mathbf{1}_{n_r}'$ (Lee and Yu, 2010). Let $\left(F_{n_r}, \frac{1}{\sqrt{n_r}} \mathbf{1}_{n_r} \right)$ be the matrix containing the orthonormal eigenvectors of J_r, where F_{n_r} is the

$n_r \times (n_r - 1)$ matrix that contains the eigenvectors corresponding to the eigenvalues of one.[1] Pre-multiplying both sides of (3.1) with $F_{n_r}^{'}$, we obtain,

$$F_{n_r}^{'} e^{\alpha W_r} Y_r = Z_r^* \beta + U_r^*, \quad F_{n_r}^{'} e^{\tau M_r} U_r = V_r^* \tag{3.2}$$

where $Z_r^* = F_{n_r}^{'} Z_r$, $U_r^* = F_{n_r}^{'} U_r$ and $V_r^* = F_{n_r}^{'} V_r$. Under Assumption 2, we have

$$\begin{aligned} F_{n_r}^{'} e^{\alpha W_r} Y_r &= F_{n_r}^{'} e^{\alpha W_r} \left(F_{n_r} F_{n_r}^{'} + \frac{1}{n_r} \mathbf{1}_{n_r} \mathbf{1}_{n_r}^{'} \right) Y_r \\ &= F_{n_r}^{'} e^{\alpha W_r} F_{n_r} F_{n_r}^{'} Y_r + \frac{1}{n_r} F_{n_r}^{'} e^{\alpha W_r} \mathbf{1}_{n_r} \mathbf{1}_{n_r}^{'} Y_r \\ &= F_{n_r}^{'} e^{\alpha W_r} F_{n_r} Y_r^* + \frac{1}{n_r} F_{n_r}^{'} e^{\alpha W_r} \mathbf{1}_{n_r} \mathbf{1}_{n_r}^{'} Y_r, \end{aligned}$$

where $Y_r^* = F_{n_r}^{'} Y_r$. Note that $F_{n_r}^{'} e^{\alpha W_r} \mathbf{1}_{n_r} \mathbf{1}_{n_r}^{'} Y_r = \sum_{i=0}^{\infty} \frac{\alpha^i}{i!} F_{n_r}^{'} W_r^i \mathbf{1}_{n_r} \mathbf{1}_{n_r}^{'} Y_r = 0$ because $W_r \mathbf{1}_{n_r} = \mathbf{1}_{n_r}$ and $F_{n_r}^{'} \mathbf{1}_{n_r} = 0$. Also, note that

$$F_{n_r}^{'} e^{\alpha W_r} F_{n_r} = \sum_{i=0}^{\infty} \frac{\alpha^i}{i!} F_{n_r}^{'} W_r^i F_{n_r} = \sum_{i=0}^{\infty} \frac{\alpha^i}{i!} \left(F_{n_r}^{'} W_r F_{n_r} \right)^i = e^{\alpha F_{n_r}^{'} W F_{n_r}}.$$

Thus, we have $F_{n_r}^{'} e^{\alpha W_r} Y_r = e^{\alpha W_r^*} Y_r^*$, where $W_r^* = F_{n_r}^{'} W_r F_{n_r}$. Similarly, we can write $F_{n_r}^{'} e^{\tau M_r} U_r = e^{\tau M_r^*} U_r^*$, where $M_r^* = F_{n_r}^{'} M_r F_{n_r}$. Then, (3.2) can be written as

$$e^{\alpha W_r^*} Y_r^* = Z_r^* \beta + U_r^*, \quad e^{\tau M_r^*} U_r^* = V_r^*. \tag{3.3}$$

Note that the transformation reduces the effective number of observations in the rth group from n_r to $n_r^* = (n_r - 1)$. Let $N = \sum_{r=1}^{R} n_r^* = n - R$ be the total number of observations in the transformed model, $\theta = (\gamma^{'}, \sigma^2)^{'}$ with $\gamma = (\beta^{'}, \zeta^{'})^{'}$ and $\zeta = (\alpha, \tau)^{'}$. Then, under Assumption 1, the quasi log-likelihood function of (3.3) can be expressed as

$$\begin{aligned} \ln L(\theta) &= -\frac{N}{2} \ln\left(2\pi\sigma^2\right) + \sum_{r=1}^{R} \ln\left| e^{\alpha W_r^*} \right| + \sum_{r=1}^{R} \ln\left| e^{\tau M_r^*} \right| - \frac{1}{2\sigma^2} \sum_{r=1}^{R} V_r^{*'}(\gamma) V_r^*(\gamma) \\ &= -\frac{N}{2} \ln\left(2\pi\sigma^2\right) - R\left(\alpha + \tau\right) - \frac{1}{2\sigma^2} \sum_{r=1}^{R} V_r^{*'}(\gamma) V_r^*(\gamma), \end{aligned} \tag{3.4}$$

where $|\cdot|$ is the determinant operator and $V_r^*(\gamma) = e^{\tau M_r^*}\left(e^{\alpha W_r^*} Y_r^* - Z_r^* \beta\right)$. The second equality in (3.4) follows from the fact that $\ln\left|e^{\alpha W_r^*}\right| = \ln e^{\alpha \operatorname{tr}(W_r^*)} = -\alpha$ and $\ln\left|e^{\tau M_r^*}\right| =$

[1] Some properties of F_{n_r} are (i) $J_{n_r} F_{n_r} = F_{n_r}$, (ii) $F_{n_r}^{'} F_{n_r} = I_{n_r - 1}$, (iii) $F_{n_r}^{'} \mathbf{1}_{n_r} = 0$, (iv) $F_{n_r} F_{n_r}^{'} + \frac{1}{n_r} \mathbf{1}_{n_r} \mathbf{1}_{n_r}^{'} = I_{n_r}$, and (v) $F_{n_r} F_{n_r}^{'} = J_{n_r}$.

$\ln e^{\tau \mathrm{tr}(M_r^*)} = -\tau$, where $\mathrm{tr}(\cdot)$ is the trace operator. Thus, $\ln L(\theta)$ is free of any Jacobian terms that need to be computed at each iteration during the estimation process.

Using $J_r = F_{n_r} F_{n_r}^{'}$ and the fact that $F_{n_r}^{'} e^{\tau M_r} J_r = F_{n_r}^{'} e^{\tau M_r}$, we can express $V_r^*(\gamma)$ as

$$\begin{aligned} V_r^*(\gamma) &= e^{\tau M_r^*}\left(e^{\alpha W_r^*} Y_r^* - Z_r^* \beta\right) = F_{n_r}^{'} e^{\tau M_r} F_{n_r}\left(F_{n_r}^{'} e^{\alpha W_r} Y_r - F_{n_r}^{'} Z_r \beta\right) \\ &= F_{n_r}^{'} e^{\tau M_r} J_r \left(e^{\alpha W_r} Y_r - Z_r \beta\right) = F_{n_r}^{'} e^{\tau M_r}\left(e^{\alpha W_r} Y_r - Z_r \beta\right) = F_{n_r}^{'} V_r(\gamma), \end{aligned} \tag{3.5}$$

where $V_r(\gamma) = e^{\tau M_r}\left(e^{\alpha W_r} Y_r - Z_r \beta\right)$. Thus, we have $V_r^{*'}(\gamma) V_r^*(\gamma) = V_r^{'}(\gamma) F_{n_r} F_{n_r}^{'} V_r(\gamma) = V_r^{'}(\gamma) J_r V_r(\gamma)$. This important property allows us to express $\ln L(\theta)$ in terms of the original variables:

$$\ln L(\theta) = -\frac{N}{2} \ln\left(2\pi\sigma^2\right) - R(\alpha + \tau) - \frac{1}{2\sigma^2} \sum_{r=1}^{R} V_r^{'}(\gamma) J_r V_r(\gamma). \tag{3.6}$$

Let $\mathrm{Blkdiag}(A_1, \ldots, A_R)$ be the block-diagonal matrix formed by constant matrices $A_1, \ldots, A_R$. Define $e^{\alpha W} = \mathrm{Blkdiag}(e^{\alpha W_1}, \ldots, e^{\alpha W_R})$, $e^{\tau M} = \mathrm{Blkdiag}(e^{\tau M_1}, \ldots, e^{\tau M_R})$, $W = \mathrm{Blkdiag}(W_1, \ldots, W_R)$, $M = \mathrm{Blkdiag}(M_1, \ldots, M_R)$, $J_n = \mathrm{Blkdiag}(J_1, \ldots, J_R)$, $Y = (Y_1^{'}, \ldots, Y_R^{'})^{'}$, $Z = (Z_1^{'}, \ldots, Z_R^{'})^{'}$, $U = (U_1^{'}, \ldots, U_R^{'})^{'}$ and $V = (V_1^{'}, \ldots, V_R^{'})^{'}$. Then, (3.6) can be expressed as

$$\ln L(\theta) = -\frac{N}{2} \ln\left(2\pi\sigma^2\right) - R(\alpha + \tau) - \frac{1}{2\sigma^2} V^{'}(\gamma) J_n V(\gamma), \tag{3.7}$$

where $V(\gamma) = e^{\tau M}(e^{\alpha W} Y - Z\beta)$. From the first-order conditions with respect to β and σ^2, we respectively have

$$\widehat{\beta}(\zeta) = \left(Z^{'} e^{\tau M^{'}} J_n e^{\tau M} Z\right)^{-1}\left(Z^{'} e^{\tau M^{'}} J_n e^{\tau M} e^{\alpha W} Y\right), \tag{3.8}$$

$$\widehat{\sigma}^2(\zeta) = \frac{1}{N}\left(e^{\alpha W} Y - Z\widehat{\beta}(\zeta)\right)^{'} e^{\tau M^{'}} J_n e^{\tau M}\left(e^{\alpha W} Y - Z\widehat{\beta}(\zeta)\right). \tag{3.9}$$

Substituting (3.8) and (3.9) into (3.7), we get the concentrated quasi log-likelihood function,

$$\ln L(\zeta) = -\frac{N}{2}\left(\ln(2\pi) + 1\right) - R(\alpha + \tau) - \frac{N}{2} \ln \widehat{\sigma}^2(\zeta). \tag{3.10}$$

Let $\theta_0 = (\gamma_0^{'}, \sigma_0^2)^{'}$ with $\gamma_0 = (\beta_0^{'}, \zeta_0^{'})^{'}$ be the true parameter vector. Then, the QML estimator $\widehat{\zeta}$ of ζ_0 maximizes (3.10), i.e., $\widehat{\zeta} = \mathrm{argmax}_{\zeta} \ln L(\zeta)$. Upon substituting $\widehat{\zeta}$ into (3.8) and (3.9), we obtain the QML estimators $\widehat{\beta} = \widehat{\beta}(\widehat{\zeta})$ and $\widehat{\sigma}^2 = \widehat{\sigma}^2(\widehat{\zeta})$ of β_0 and σ_0^2, respectively.

To ensure that ζ_0 is globally identified in our model, we require that $\frac{1}{N}(G(\zeta) - G(\zeta_0)) > 0$ for any $\zeta \neq \zeta_0$, where $G(\zeta) = \max_{\beta, \sigma^2} \mathrm{E}\left(\ln L(\zeta)\right)$. To investigate this condition, we define the following terms:

$$K_{11}(\zeta) = Z'e^{\tau M'}J_n e^{\tau M}Z, \quad K_{12}(\zeta) = K_{21}'(\zeta) = Z'e^{\tau M'}J_n e^{\tau M}e^{(\alpha-\alpha_0)W}Z\beta_0,$$
$$K_{22}(\zeta) = (e^{(\alpha-\alpha_0)W}Z\beta_0)'e^{\tau M'}J_n e^{\tau M}e^{(\alpha-\alpha_0)W}Z\beta_0. \tag{3.11}$$

Assumption 5. *Z is exogenous, with uniformly bounded elements, and has full column rank. Also $\lim_{N\to\infty}\frac{1}{N}Z'e^{\tau M'}J_n e^{\tau M}Z$ exists and is non-singular, uniformly in $\tau \in [-c, c]$.*

Assumption 6. *Either*

(a) $\lim_{N\to\infty}\frac{1}{N}\overline{K}(\zeta) = \lim_{N\to\infty}\frac{1}{N}\left(K_{22}(\zeta) - K_{12}'(\zeta)K_{11}^{-1}(\zeta)K_{12}(\zeta)\right) \geq 0$, and the equality holds only when $\zeta = \zeta_0$; or

(b) $\lim_{N\to\infty}\ln(\sigma^2(\zeta)) - \ln(\sigma_0^2) + 2R(\alpha-\alpha_0) + 2R(\tau-\tau_0)$, where $\sigma^2(\zeta) = \frac{\sigma_0^2}{N}\mathrm{tr}(A'(\zeta)J_n A(\zeta))$ with $A(\zeta) = e^{\tau M}e^{(\alpha-\alpha_0)W}e^{-\tau_0 M}$, is not zero for $\zeta \neq \zeta_0$.

Assumption 5 provides some regularity conditions and corresponds to Assumption 4.2 in Lee et al. (2010). Assumption 6 provides sufficient conditions for the global identification of ζ_0 and can be considered as the extension of Assumption 5.1 in Lee et al. (2010) to our setting (see the proof of Theorem 3.1 for the details). The following theorem states the consistency of the QML estimator.

Theorem 3.1. *Under Assumptions 1–6, $\widehat{\theta}$ is a consistent estimator of θ_0, i.e., $\widehat{\theta} \xrightarrow{p} \theta_0$.*

Proof. See Section B.1 in the web appendix. □

To derive the asymptotic distribution of $\widehat{\theta}$, we apply the mean value theorem to the score functions to get $\sqrt{N}(\widehat{\theta}-\theta_0) = -\left(\frac{1}{N}\frac{\partial^2 \ln L(\overline{\theta})}{\partial\theta\partial\theta'}\right)^{-1}\frac{1}{\sqrt{N}}\frac{\partial \ln L(\theta_0)}{\partial\theta}$, where $\overline{\theta}$ lies between $\widehat{\theta}$ and θ_0 element-wise (Jennrich, 1969, Lemma 3). Let $\Psi(\theta_0) = -\frac{1}{N}\mathrm{E}\left(\frac{\partial^2 \ln L(\theta_0)}{\partial\theta\partial\theta'}\right)$, which can be derived as

$$\Psi(\theta_0) = \frac{1}{N\sigma_0^2}\begin{pmatrix} K_{11}^{\star} & * & * & * \\ K_{21}^{\star} & K_{22}^{\star} & * & * \\ 0_{1\times k} & 0 & 0 & * \\ 0_{1\times k} & 0 & 0 & 0 \end{pmatrix} + \frac{1}{N}\begin{pmatrix} 0_{k\times k} & * & * & * \\ 0_{1\times k} & \mathrm{tr}(J_n\mathbb{W}(J_n\mathbb{W})^s) & * & * \\ 0_{1\times k} & \mathrm{tr}(J_n\mathbb{W}(J_n M)^s) & \mathrm{tr}((J_n M)(J_n M)^s) & * \\ 0_{1\times k} & \frac{R}{\sigma_0^2} & \frac{R}{\sigma_0^2} & \frac{N}{2\sigma_0^4} \end{pmatrix},$$

where $K_{11}^{\star} = Z'e^{\tau_0 M'}J_n e^{\tau_0 M}Z$, $K_{21}^{\star} = -\beta_0'Z'W'e^{\tau_0 M'}J_n e^{\tau_0 M}Z$, $K_{22}^{\star} = \beta_0'Z'W'e^{\tau_0 M'}J_n e^{\tau_0 M}WZ\beta_0$, $\mathbb{W} = e^{\tau_0 M}We^{-\tau_0 M}$, and $A^s = A + A'$ for any square matrix A. It can be shown that the limit of $\Psi(\theta_0)$ is non-singular under the following assumption (see the proof of Theorem 3.2 for the details).

Assumption 7. $\lim_{N\to\infty}\frac{1}{N}\left(\frac{\mathrm{tr}((J_n\mathbb{W})(J_n\mathbb{W})^s)\mathrm{tr}((J_nM)(J_nM)^s)-\mathrm{tr}^2((J_n\mathbb{W})(J_nM^s))}{\mathrm{tr}((J_nM)(J_nM)^s)}\right)$ *is strictly positive.*

Let $\mu_3=\mathrm{E}(v_{ir}^3)$ and $\mu_4=\mathrm{E}(v_{ir}^4)$, and define $\kappa=(\mu_4-3\sigma_0^4)/\sigma_0^4$. Then, we can show that $\Omega(\theta_0)=\mathrm{Var}\left(\frac{1}{\sqrt{N}}\frac{\partial \ln L(\theta_0)}{\partial\theta}\right)=\Psi(\theta_0)+\Upsilon$, where

$$\Upsilon=\frac{1}{N}\begin{pmatrix} 0_{k\times k} & * & * & * \\ -\frac{\mu_3}{\sigma_0^4}\mathrm{vec}_D'(J_n\mathbb{W})J_ne^{\tau_0M}Z & \Upsilon_{22} & * & * \\ -\frac{\mu_3}{\sigma_0^4}\mathrm{vec}_D'(J_nM)J_ne^{\tau_0M}Z & \Upsilon_{32} & \kappa\mathrm{vec}_D'(J_nM)\mathrm{vec}_D(J_nM) & * \\ 0_{1\times k} & \Upsilon_{42} & -\frac{\kappa}{2\sigma_0^2}\mathrm{vec}_D'(J_nM)\mathrm{vec}_D(J_n) & \frac{\kappa}{4\sigma_0^4}\mathrm{vec}_D'(J_n)\mathrm{vec}_D(J_n)\end{pmatrix},$$

and

$$\begin{aligned}\Upsilon_{22}&=\frac{2\mu_3}{\sigma_0^4}\mathrm{vec}_D'(J_n\mathbb{W})J_ne^{\tau_0M}WZ\beta_0+\kappa\mathrm{vec}_D'(J_n\mathbb{W})\mathrm{vec}_D(J_n\mathbb{W}),\\ \Upsilon_{32}&=\frac{\mu_3}{\sigma_0^4}\mathrm{vec}_D'(J_nM)J_ne^{\tau_0M}WZ\beta_0+\kappa\mathrm{vec}_D'(J_n\mathbb{W})\mathrm{vec}_D(J_nM),\\ \Upsilon_{42}&=-\frac{\kappa}{2\sigma_0^2}\mathrm{vec}_D'(J_n\mathbb{W})\mathrm{vec}_D(J_n).\end{aligned}$$

Theorem 3.2. *Under Assumptions 1–7, we have*

$$\sqrt{N}(\widehat{\theta}-\theta_0)\xrightarrow{d}N\left(0,\lim_{N\to\infty}\Psi^{-1}(\theta_0)\Omega(\theta_0)\Psi^{-1}(\theta_0)\right),$$

where $\Psi(\theta_0)=-\frac{1}{N}\mathrm{E}\left(\frac{\partial^2\ln L(\theta_0)}{\partial\theta\partial\theta'}\right)$ *and* $\Omega(\theta_0)=\mathrm{Var}\left(\frac{1}{\sqrt{N}}\frac{\partial\ln L(\theta_0)}{\partial\theta}\right)$.

Proof. See Section B.2 in the web appendix. □

For inference, we suggest using the plug-in estimators of $\Psi(\theta_0)$ and $\Omega(\theta_0)$ formulated with $\widehat{\theta}$. The expression for $\Omega(\theta_0)$ contains μ_3 and κ. To define consistent estimators of μ_3 and κ, let $\tilde{V}=J_nV$ and f_{jh} be the (j,h)th element of J_n for $j,h=1,\ldots,n$. Also, let $\tilde{v}_j$ and v_j be the jth element of $\tilde{V}$ and V, respectively, for $j=1,2,\ldots,n$. Then, $\tilde{v}_j=f_{j1}v_1+\ldots+f_{jn}v_n$, for $j=1,2,\ldots,n$. Assumption 1 ensures that

$$\mathrm{E}(\tilde{v}_j^4)=\sum_{h=1}^n f_{jh}^4\mathrm{E}(v_h^4)+3\sigma_0^4\sum_{h=1}^n\sum_{l=1}^n f_{jh}^2f_{jl}^2-3\sigma_0^4\sum_{h=1}^n f_{jh}^4=\kappa\sum_{h=1}^n f_{jh}^4\sigma_0^4+3\sigma_0^4\sum_{h=1}^n\sum_{l=1}^n f_{jh}^2f_{jl}^2.$$

Summing $\mathrm{E}(\tilde{v}_j^4)$ over j and solving for κ, we obtain

$$\kappa=\frac{\sum_{j=1}^n\mathrm{E}(\tilde{v}_j^4)-3\sigma_0^4\sum_{j=1}^n\sum_{h=1}^n\sum_{l=1}^n f_{jh}^2f_{jl}^2}{\sigma_0^4\sum_{j=1}^n\sum_{h=1}^n f_{jh}^4}.$$

This last expression suggests that a consistent estimator for κ can be formulated as

$$\widehat{\kappa}=\frac{\sum_{j=1}^{n}\widehat{\tilde{v}}_j^4-3\widehat{\sigma}^4\sum_{j=1}^{n}\sum_{h=1}^{n}\sum_{l=1}^{n}f_{jh}^2f_{jl}^2}{\widehat{\sigma}^4\sum_{j=1}^{N}\sum_{h=1}^{n}f_{jh}^4} \tag{3.12}$$

where $\widehat{\sigma}^2=\widehat{\sigma}^2(\widehat{\zeta})$ and $\widehat{\tilde{v}}_j$ is the jth element of $J_n e^{\widehat{\tau}M}\left(e^{\widehat{\alpha}W}Y-Z\widehat{\beta}\right)$. Similarly, under Assumption 1, we have $\mathrm{E}(\tilde{v}_j^3)=\sum_{h=1}^{n}f_{jh}^3\mathrm{E}(v_h^3)=\mu_3\sum_{h=1}^{n}f_{jh}^3$, which gives $\mu_3=\mathrm{E}(\tilde{v}_j^3)/\sum_{h=1}^{n}f_{jh}^3$. This last expression suggests the following estimator for μ_3:

$$\widehat{\mu}_3=\frac{\sum_{j=1}^{n}\widehat{\tilde{v}}_j^3}{\sum_{j=1}^{n}\sum_{h=1}^{n}f_{jh}^3}. \tag{3.13}$$

3.2 Estimation under heteroskedasticity

In this section, we consider the estimation of the MES network model under heteroskedasticity. The estimation approach consists of two steps. In the first step, the score functions obtained from the quasi log-likelihood function are adjusted so that their expectations at the true parameter vector are zero under heteroskedasticity. In the second step, we define our suggested estimator as the root of the adjusted score functions. We start by specifying the form of heteroskedasticity.

Assumption 8. *The v_{ir}'s are independently distributed over r and i with $\mathrm{E}(v_{ir})=0$ and $\mathrm{Var}(v_{ir})=\sigma_{ir}^2$, and $\mathrm{E}|v_{ir}|^{4+\varrho}<\infty$ for some $\varrho>0$.*

We will denote the true parameter value as $\theta_0=(\beta_0',\zeta_0')'$ with $\zeta_0=(\alpha_0,\tau_0)'$, and any arbitrary value in the parameter space as $\theta=(\beta',\zeta')'$ with $\zeta=(\alpha,\tau)'$. Based on (3.7), the relevant quasi-score functions are

$$S(\theta)=\begin{cases}\beta:\ Z'e^{\tau M'}J_nV(\theta),\\ \alpha:\ -V'(\theta)J_nY(\zeta)-R,\\ \tau:\ -V'(\theta)J_nMV(\theta)-R,\end{cases} \tag{3.14}$$

where $V(\theta)=e^{\tau M}\left(e^{\alpha W}Y-Z\beta\right)$ and $Y(\zeta)=e^{\tau M}We^{\alpha W}Y$. Let $\Sigma=\mathrm{Blkdiag}(\Sigma_1,\ldots,\Sigma_R)$, where $\Sigma_r=\mathrm{Diag}(\sigma_{1r}^2,\ldots,\sigma_{n_r r}^2)$ is the $n_r\times n_r$ diagonal matrix containing the unknown variance terms for $r = 1, \ldots, R$. Then, $\mathrm{E}\left(S(\theta_0)\right)$ can be derived as

$$\mathrm{E}(S(\theta_0)) = \begin{cases} \beta: \; \mathrm{E}\left(Z' e^{\tau_0 M'} J_n V\right) = 0, \\ \alpha: \; -\mathrm{E}\left(V' J_n Y(\zeta_0)\right) - R = -\mathrm{tr}(\Sigma J_n \mathbb{W}) - R, \\ \tau: \; -\mathrm{E}\left(V' J_n M V\right) - R = -\mathrm{tr}(\Sigma J_n M) - R, \end{cases} \tag{3.15}$$

where $\mathbb{W} = e^{\tau_0 M} W e^{-\tau_0 M}$. Under our assumptions, it follows that $\mathrm{tr}(\Sigma J_n \mathbb{W}) = O(N)$ and $\mathrm{tr}(\Sigma J_n M) = O(R)$. Thus, $\mathrm{plim}_{N\to\infty} \frac{1}{N} S(\theta_0) \neq 0$ because $\frac{1}{N}\mathrm{E}(S(\theta_0)) = O(1)$. This observation suggests that the QMLE may not be consistent under heteroskedasticity. However, if W and M are commutative, i.e., $WM = MW$, then $\mathbb{W} = e^{\tau_0 M} W e^{-\tau_0 M} = W$. Thus, we have $\mathrm{tr}(\Sigma J_n \mathbb{W}) = \mathrm{tr}\left(\Sigma J_n W\right) = \sum_{r=1}^{R} \mathrm{tr}\left(\Sigma_r J_r W_r\right) = -\sum_{r=1}^{R} \frac{1}{n_r} \mathrm{tr}(\Sigma_r \mathbf{1}_{n_r} \mathbf{1}_{n_r}' W_r) = O(R)$. Hence, $\mathrm{plim}_{N\to\infty} \frac{1}{N} S(\theta_0) = 0$ when $R/N \to 0$.

We will adjust the quasi-score functions such that their expectations become zero in all cases. We start with the score function with respect to α. First, note that $\mathrm{tr}(DA) = \mathrm{tr}(D\,\mathrm{Diag}(A))$ for a diagonal matrix D and any conformable matrix A. Using this property and $e^{\alpha_0 W} Y = Z\beta_0 + C_\lambda \lambda_0 + e^{-\tau_0 M} V$, where $C_\lambda = \mathrm{Blkdiag}(\mathbf{1}_{n_1}, \ldots, \mathbf{1}_{n_R})$ and $\lambda_0 = (\lambda_1, \ldots, \lambda_R)'$, we have

$$\begin{aligned} &\mathrm{E}\left(V' J_n Y(\zeta_0)\right) = \mathrm{E}\left(V' J_n e^{\tau_0 M} W e^{\alpha_0 W} Y\right) = \mathrm{E}\left(V' J_n e^{\tau_0 M} W e^{-\tau_0 M} e^{\tau_0 M} e^{\alpha_0 W} Y\right) \\ &= \mathrm{E}\left(V' J_n \mathbb{W} e^{\tau_0 M} e^{\alpha_0 W} Y\right) = \mathrm{E}\left(V' J_n \mathbb{W} V\right) = \mathrm{tr}\left(\Sigma J_n \mathbb{W}\right) = \mathrm{tr}\left(\Sigma\,\mathrm{Diag}(J_n \mathbb{W})\right) \\ &= \mathrm{tr}\left(\Sigma\,\mathrm{Diag}(J_n \mathbb{W})(\mathrm{Diag}(J_n))^{-1} J_n\right) = \mathrm{E}\left(V'\,\mathrm{Diag}(J_n \mathbb{W})(\mathrm{Diag}(J_n))^{-1} J_n V\right). \end{aligned}$$

Then, subtracting the last term from the fifth term, we obtain

$$\mathrm{E}\left(V' J_n \mathbb{W} V\right) - \mathrm{E}\left(V'\,\mathrm{Diag}(J_n \mathbb{W})(\mathrm{Diag}(J_n))^{-1} J_n V\right) = 0 \implies \mathrm{E}\left(V' \mathbb{W}_D V\right) = 0, \tag{3.16}$$

where $\mathbb{W}_D = J_n \mathbb{W} - \mathrm{Diag}(J_n \mathbb{W})(\mathrm{Diag}(J_n))^{-1} J_n$. Similarly, with respect to the τ element of $\mathrm{E}\left(S(\theta_0)\right)$, we have

$$\begin{aligned} \mathrm{E}\left(V' J_n M V\right) &= \mathrm{tr}\left(\Sigma J_n M\right) = \mathrm{tr}\left(\Sigma\,\mathrm{Diag}(J_n M)\right) = \mathrm{tr}\left(\Sigma\,\mathrm{Diag}(J_n M)(\mathrm{Diag}(J_n))^{-1} J_n\right) \\ &= \mathrm{E}\left(V'\,\mathrm{Diag}(J_n M)(\mathrm{Diag}(J_n))^{-1} J_n V\right). \end{aligned}$$

Then, subtracting the last term from the first term, we obtain

$$\mathrm{E}\left(V^{'}\mathbb{M}_D V\right) = 0, \tag{3.17}$$

where $\mathbb{M}_D = J_n M - \mathrm{Diag}(J_n M)(\mathrm{Diag}(J_n))^{-1} J_n$. Hence, we propose the following adjusted quasi score functions for estimation:

$$S^*(\theta) = \begin{cases} \beta : \; Z^{'} e^{\tau M^{'}} J_n V(\theta), \\ \alpha : \; -V^{'}(\theta)\mathbb{W}_D(\tau)V(\theta), \\ \tau : \; -V^{'}(\theta)\mathbb{M}_D V(\theta), \end{cases} \tag{3.18}$$

where $\mathbb{W}_D(\tau) = J_n\mathbb{W}^*(\tau) - \mathrm{Diag}(J_n\mathbb{W}^*(\tau))(\mathrm{Diag}(J_n))^{-1}J_n$, $\mathbb{W}^*(\tau) = \mathbb{W}(\tau) - \mathrm{Diag}(\mathbb{W}(\tau))$, and $\mathbb{W}(\tau) = e^{\tau M} W e^{-\tau M}$. The robust M-estimator can thus be derived by solving the system of nonlinear equations in (3.18). The solution can be simplified by first solving for β for a given ζ value, which can then be substituted into (3.18) to obtain the concentrated quasi score functions with respect to ζ. By setting the quasi score function with respect to β to zero, we obtain the following estimator for a given ζ value:

$$\widehat{\beta}_M(\zeta) = \left(Z^{'} e^{\tau M^{'}} J_n e^{\tau M} Z\right)^{-1} \left(Z^{'} e^{\tau M^{'}} J_n e^{\tau M} e^{\alpha W} Y\right). \tag{3.19}$$

Substituting $\widehat{\beta}_M(\zeta)$ back into the α and τ elements of $S^*(\theta)$, we obtain the following concentrated quasi score functions:

$$S^{*c}(\zeta) = \begin{cases} \alpha : \; -\widehat{V}^{'}(\zeta)\mathbb{W}_D(\tau)\widehat{V}(\zeta), \\ \tau : \; -\widehat{V}^{'}(\zeta)\mathbb{M}_D\widehat{V}(\zeta), \end{cases} \tag{3.20}$$

where $\widehat{V}(\zeta) = V(\widehat{\beta}_M(\zeta), \zeta) = e^{\tau M}\left(e^{\alpha W}Y - Z\widehat{\beta}_M(\zeta)\right)$. Then, we define the robust M-estimator (RME) $\widehat{\zeta}_M$ of ζ_0 by $\widehat{\zeta}_M = \mathrm{argsolve}\{S^{*c}(\zeta) = 0\}$. Substituting $\widehat{\zeta}_M$ into (3.19) yields the RME $\widehat{\beta}_M = \widehat{\beta}_M(\widehat{\zeta}_M)$ of β_0. Note that the consistency of $\widehat{\theta}_M = (\widehat{\beta}_M^{'}, \widehat{\zeta}_M^{'})^{'}$ follows from the consistency of $\widehat{\zeta}_M$, since $\widehat{\beta}_M = \widehat{\beta}_M(\widehat{\zeta}_M)$. To prove the consistency of $\widehat{\zeta}_M$, we define the population counterpart of (3.18) as

$$\bar{S}^*(\theta) = \mathrm{E}(S^*(\theta)) = \begin{cases} \beta : \; \mathrm{E}\left(Z^{'} e^{\tau M^{'}} J_n V(\theta)\right), \\ \alpha : \; -\mathrm{E}\left(V^{'}(\theta)\mathbb{W}_D(\tau)V(\theta)\right), \\ \tau : \; -\mathrm{E}\left(V^{'}(\theta)\mathbb{M}_D V(\theta)\right). \end{cases} \tag{3.21}$$

Setting the score function with respect to β to zero yields

$$\overline{\beta}_M(\zeta) = \left(Z' e^{\tau M'} J_n e^{\tau M} Z\right)^{-1} \left(Z' e^{\tau M'} J_n e^{\tau M} e^{\alpha W} \mathrm{E}(Y)\right). \tag{3.22}$$

Substituting this expression into the α and τ elements of (3.21), we obtain the population counterparts of (3.20):

$$\overline{S}^{*c}(\zeta) = \begin{cases} \alpha : -\mathrm{E}\left(\overline{V}'(\zeta)\mathbb{W}_D(\tau)\overline{V}(\zeta)\right), \\ \tau : -\mathrm{E}\left(\overline{V}'(\zeta)\mathbb{M}_D\overline{V}(\zeta)\right), \end{cases} \tag{3.23}$$

where $\overline{V}(\zeta) = V(\overline{\beta}_M(\zeta), \zeta)$. To prove the consistency of $\widehat{\zeta}_M$, we need to show (i) the uniform convergence of $S^{*c}(\zeta)$ to $\overline{S}^{*c}(\zeta)$ over $\zeta \in \Delta$, i.e., $\sup_{\zeta\in\Delta} \frac{1}{N}\left\|S^{*c}(\zeta) - \overline{S}^{*c}(\zeta)\right\| = o_p(1)$, and (ii) the identification-uniqueness condition (van der Vaart, 1998, Theorem 5.9). The following high-level assumption states the identification-uniqueness condition for ζ_0[2].

Assumption 9. $\inf_{\zeta: d(\zeta,\zeta_0)\geq\nu} \left\|\overline{S}^{*c}(\zeta)\right\| > 0$ *for every* $\nu > 0$*, where* $d(\zeta, \zeta_0)$ *is a measure of distance between* ζ *and* ζ_0*.*

Theorem 3.3. *Under Assumptions 2–5 and 8–9, we have* $\widehat{\theta}_M \xrightarrow{p} \theta_0$ *as* $n \longrightarrow \infty$*.*

Proof. See Section B.3 in the web appendix. □

To derive the asymptotic distribution of the RME, we apply the mean value theorem to $S^*(\widehat{\theta}_M)$ at θ_0, to obtain $\sqrt{N}(\widehat{\theta}_M - \theta_0) = -\left(\frac{1}{N}\frac{\partial S^*(\overline{\theta})}{\partial\theta'}\right)^{-1} \frac{1}{\sqrt{N}}S^*(\theta_0)$, where $\overline{\theta}$ lies between θ_0 and $\widehat{\theta}_M$ element-wise. In order to apply the CLT to $\frac{1}{\sqrt{N}}S^*(\theta_0)$, we need to show that $S^*(\theta_0)$ can be written in terms of linear and quadratic forms of V (see Lemma A.4 in the web appendix). Since $J_nV(\theta_0) = J_nV$, the adjusted score function in $S^*(\theta_0)$ with respect to β can be expressed as $Z'e^{\tau_0 M'}J_nV(\theta_0) = Z'e^{\tau_0 M'}J_nV$. For the α element of $S^*(\theta_0)$, we have $V'(\theta_0)\mathbb{W}_DV'(\theta_0)$, where $V(\theta_0) = e^{\tau_0 M}\left(e^{\alpha_0 W}Y - Z\beta_0\right) = e^{\tau_0 M}C_\lambda\lambda_0 + V$. Then,

$$\begin{aligned}
\frac{1}{\sqrt{N}}V'(\theta_0)\mathbb{W}_DV(\theta_0) &= \frac{1}{\sqrt{N}}(e^{\tau_0 M}C_\lambda\lambda_0 + V)'\mathbb{W}_D\left(e^{\tau_0 M}C_\lambda\lambda_0 + V\right) \\
&= \frac{1}{\sqrt{N}}V'\mathbb{W}_DV + \frac{1}{\sqrt{N}}V'\mathbb{W}_De^{\tau_0 M}C_\lambda\lambda_0 + \frac{1}{\sqrt{N}}\lambda_0'C_\lambda'e^{\tau_0 M'}\mathbb{W}_DV \\
&\quad + \frac{1}{\sqrt{N}}\lambda_0'C_\lambda'e^{\tau_0 M'}\mathbb{W}_De^{\tau_0 M}C_\lambda\lambda_0 = \frac{1}{\sqrt{N}}V'\mathbb{W}_DV + o_p(1),
\end{aligned}$$

[2] In Section B.6 in the web appendix, we provide some low-level conditions that are sufficient for Assumption 9 to hold.

because $\mathbb{W}_D e^{\tau_0 M} C_\lambda \lambda_0 = 0$ and $\frac{1}{\sqrt{N}} \lambda_0' C_\lambda' e^{\tau_0 M'} \mathbb{W}_D V = o_p(1)$ by Lemma A.6 in the web appendix. Similarly, we have $\frac{1}{\sqrt{N}} V(\theta_0) \mathbb{M}_D V(\theta_0) = \frac{1}{\sqrt{N}} V' \mathbb{M}_D V + o_p(1)$ by Lemma A.6 in the web appendix. Thus, we can express $S^*(\theta_0)$ in terms of linear and quadratic forms of V as

$$\frac{1}{\sqrt{N}} S^*(\theta_0) = \begin{cases} \beta: \ \frac{1}{\sqrt{N}} Z' e^{\tau_0 M'} J_n V, \\ \alpha: \ -\frac{1}{\sqrt{N}} V' \mathbb{W}_D V + o_p(1), \\ \tau: \ -\frac{1}{\sqrt{N}} V' \mathbb{M}_D V + o_p(1). \end{cases} \tag{3.24}$$

This result in (3.24), combined with $\frac{1}{N} \frac{\partial S^*(\bar{\theta})}{\partial \theta'} - \frac{1}{N} \mathrm{E}\left(\frac{\partial S^*(\theta_0)}{\partial \theta'} \right) = o_p(1)$, leads to the following theorem.

Theorem 3.4. *Under Assumptions 2–5 and 8–9, we have*

$$\sqrt{N}(\widehat{\theta}_M - \theta_0) \xrightarrow{d} N\left(0, \lim_{N \to \infty} \Psi^{*-1}(\theta_0) \Omega^*(\theta_0) \Psi^{*-1'}(\theta_0)\right), \tag{3.25}$$

where $\Psi^*(\theta_0) = -\frac{1}{N} \mathrm{E}\left(\frac{\partial S^*(\theta_0)}{\partial \theta'} \right)$ *and* $\Omega^*(\theta_0) = \mathrm{Var}\left(\frac{1}{\sqrt{N}} S^*(\theta_0) \right)$ *are assumed to exist and* $\Psi^*(\theta_0)$ *is assumed to be positive definite for sufficiently large* N.

Proof. See Section B.4 in the web appendix. □

Theorem 3.4 indicates that we need consistent estimators of $\Psi^*(\theta_0)$ and $\Omega^*(\theta_0)$ for inference. In the case of $\Psi^*(\theta_0)$, we can use the observed counterpart given by $\Psi^*(\widehat{\theta}_M) = -\frac{1}{N} \frac{\partial S^*(\theta)}{\partial \theta'}|_{\theta = \widehat{\theta}_M}$. The elements of $\Psi^*(\theta)$ are given below.

$$\begin{aligned} &N\Psi^*_{\beta\beta}(\theta) = Z' e^{\tau M'} J_n e^{\tau M} Z, \ N\Psi^*_{\beta\alpha}(\theta) = -Z' e^{\tau M'} J_n Y(\zeta), \ N\Psi^*_{\beta\tau}(\theta) = -Z' e^{\tau M'} (J_n M)^s V(\theta), \\ &N\Psi^*_{\alpha\beta}(\theta) = -V'(\theta) \mathbb{W}^s_D(\tau) e^{\tau M} Z, N\Psi^*_{\alpha\alpha}(\theta) = Y'(\zeta) \mathbb{W}_D(\tau) V(\theta) + V'(\theta) \mathbb{W}_D(\tau) Y(\zeta), \\ &N\Psi^*_{\alpha\tau}(\theta) = V'(\theta) \left(M' \mathbb{W}_D(\tau) + \mathbb{W}_D(\tau) M + \dot{\mathbb{W}}_D(\tau) \right) V(\theta), \ N\Psi^*_{\tau\beta}(\theta) = -V'(\theta) \mathbb{M}^s_D e^{\tau M} Z \\ &N\Psi^*_{\tau\alpha}(\theta) = Y'(\zeta) \mathbb{M}^s_D V(\theta), \ N\Psi^*_{\tau\tau}(\theta) = V'(\theta) \mathbb{M}^s_D M V(\theta), \end{aligned}$$

where $\dot{\mathbb{W}}_D(\tau) = \partial \mathbb{W}_D(\tau)/\partial \tau = J_n \dot{\mathbb{W}}^*(\tau) - \mathrm{Diag}(J_n \dot{\mathbb{W}}^*(\tau)) \left(\mathrm{Diag}(J_n)\right)^{-1} J_n$ with $\dot{\mathbb{W}}^*(\tau) = \dot{\mathbb{W}}(\tau) - \mathrm{Diag}(\dot{\mathbb{W}}(\tau))$ and $\dot{\mathbb{W}}(\tau) = M\mathbb{W}(\tau) - \mathbb{W}(\tau) M$. Under our stated assumptions, it can be shown that $\Psi^*(\widehat{\theta}_M) = \Psi^*(\theta_0) + o_p(1)$.[3] In the case of $\Omega^*(\theta_0)$, we first determine the closed-form expressions of its elements by using Lemma A.2 in the web appendix:

$$\begin{aligned} &N\Omega^*_{\beta\beta}(\theta_0) = Z' e^{\tau_0 M'} J_n \Sigma J_n e^{\tau_0 M} Z, \quad N\Omega^*_{\beta\alpha}(\theta_0) = 0, \quad N\Omega^*_{\beta\tau}(\theta_0) = 0, \\ &N\Omega^*_{\alpha\alpha}(\theta_0) = \mathrm{tr}(\Sigma \mathbb{W}_D \Sigma \mathbb{W}^s_D), \ N\Omega^*_{\alpha\tau}(\theta_0) = \mathrm{tr}(\Sigma \mathbb{W}_D \Sigma \mathbb{M}^s_D), \ N\Omega^*_{\tau\tau}(\theta_0) = \mathrm{tr}(\Sigma \mathbb{M}_D \Sigma \mathbb{M}^s_D). \end{aligned} \tag{3.26}$$

[3]See the proof of Theorem 3.4 for details.

Let $\widehat{v}_{ir}$ be the ith element of $\widehat{V}_r = J_r e^{\widehat{\tau}_M M_r}\left(e^{\widehat{\alpha}_M W_r} Y_r - Z_r \widehat{\beta}_M\right)$ for $i = 1, \ldots, n_r$ and $r = 1, \ldots, R$. We replace Σ in $\Omega^*(\theta_0)$ by $\widehat{\Sigma} = \text{Diag}(\widehat{v}_{11}^2, \ldots, \widehat{v}_{n_R R}^2)$. Then, we formulate the estimator of $\Omega^*(\theta_0)$ by $\widehat{\Omega}^* = \Omega^*(\widehat{\theta}_M)$, where we replace Σ with $\widehat{\Sigma}$. In the next theorem, we show that $\widehat{\Omega}^*$ is a consistent estimator of $\Omega^*(\theta_0)$ when $R/N \to 0$, i.e., when the number of observations grows faster than the number of groups.

Theorem 3.5. *Under Assumptions 2–5 and 8–9, we have* $\widehat{\Omega}^* = \Omega^*(\theta_0) + o_p(1)$.

Proof. See Section B.5 in the web appendix. □

Remark 1. *The form of* $\widehat{\Sigma}$ *needs to be adjusted in the case of group-wise heteroskedasticity. From* $\tilde{V}_r = J_r V_r$, *we obtain* $\tilde{v}_{ir} = v_{ir} - \frac{1}{n_r}\sum_{j=1}^{n_r} v_{jr}$ *for* $r = 1, \ldots, R$. *This relation gives* $\text{E}(\tilde{v}_{ir}^2) = \frac{n_r - 1}{n_r}\sigma_r^2$ *under group-wise heteroskedasticity. Thus, we have* $\sigma_r^2 = \frac{n_r}{n_r - 1}\text{E}(\tilde{v}_{ir}^2)$. *This last result suggests that we can formulate* $\widehat{\Sigma}$ *as* $\widehat{\Sigma} = \text{Blkdiag}\left(\widehat{\sigma}_1^2 I_{n_1}, \ldots, \widehat{\sigma}_R^2 I_{n_R}\right)$, *where* $\widehat{\sigma}_r^2 = \frac{1}{n_r - 1}\sum_{i=1}^{n_r} \widehat{v}_{ir}^2$ *for* $r = 1, \ldots, R$.

4 Direct approach

In Section 3, we applied the transformation method to eliminate the group fixed effects from the model and subsequently introduced estimation for the resulting transformed model. This approach necessitates that both W and M are row-normalized as shown Assumption 2. In this section, we relax this assumption and estimate the group fixed effects along with other parameters.

4.1 Estimation under homoskedasticity

Recall that $e^{\alpha W} = \text{Blkdiag}(e^{\alpha W_1}, \ldots, e^{\alpha W_R})$, $e^{\tau M} = \text{Blkdiag}(e^{\tau M_1}, \ldots, e^{\tau M_R})$, $Y = (Y_1^{'}, \ldots, Y_R^{'})^{'}$, $Z = (Z_1^{'}, \ldots, Z_R^{'})^{'}$, $U = (U_1^{'}, \ldots, U_R^{'})^{'}$, $V = (V_1^{'}, \ldots, V_R^{'})^{'}$, $C_\lambda = \text{blkdiag}(\mathbf{1}_{n_1}, \ldots, \mathbf{1}_{n_R})$ and $\lambda = (\lambda_1, \ldots, \lambda_R)^{'}$. Then, (3.1) can be expressed in the following matrix form:

$$e^{\alpha W} Y = Z\beta + C_\lambda \lambda + U, \quad e^{\tau M} U = V. \tag{4.1}$$

Let $\theta = (\beta^{'}, \sigma^2, \zeta^{'})^{'}$ with $\zeta = (\alpha, \tau)^{'}$, and $\theta_0 = (\beta_0^{'}, \sigma_0^2, \zeta_0^{'})^{'}$ with $\zeta_0 = (\alpha_0, \tau_0)^{'}$ be the vector of true parameter values. Then, the quasi log-likelihood function can be expressed as

$$\ln L(\theta, \lambda) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} V^{'}(\beta, \zeta, \lambda) V(\beta, \zeta, \lambda), \tag{4.2}$$

where $V(\beta,\zeta,\lambda) = e^{\tau M}\left(e^{\alpha W}Y - Z\beta - C_\lambda\lambda\right)$ and $n = \sum_{r=1}^{R} n_r$. Given θ, $\ln L(\theta,\lambda)$ is partially maximized at

$$\widehat{\lambda}(\beta,\zeta) = \left(C^{'}(\tau)C(\tau)\right)^{-1} C^{'}(\tau)e^{\tau M}\left(e^{\alpha W}Y - Z\beta\right), \tag{4.3}$$

where $C(\tau) = e^{\tau M}C_\lambda$. Substituting $\widehat{\lambda}(\beta,\zeta)$ into $\ln L(\theta,\lambda)$, we obtain the following concentrated quasi log-likelihood function:

$$\ln L^c(\theta) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\tilde{V}^{'}(\beta,\zeta)\tilde{V}(\beta,\zeta), \tag{4.4}$$

where $\tilde{V}(\beta,\zeta) = Q_C(\tau)e^{\tau M}\left(e^{\alpha W}Y - Z\beta\right)$ and $Q_C(\tau) = I_n - P_C(\tau)$, with the matrix $P_C(\tau)$ being defined as $P_C(\tau) = C(\tau)\left(C^{'}(\tau)C(\tau)\right)^{-1}C^{'}(\tau)$. Based on (4.4), we derive the following score functions:

$$S^c(\theta) = \begin{cases} \beta: & \frac{1}{\sigma^2}Z^{'}e^{\tau M^{'}}\tilde{V}(\beta,\zeta), \\ \sigma^2: & -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\tilde{V}^{'}(\beta,\zeta)\tilde{V}(\beta,\zeta), \\ \alpha: & -\frac{1}{\sigma^2}Y^{'}e^{\alpha W^{'}}W^{'}e^{\tau M^{'}}\tilde{V}(\beta,\zeta), \\ \tau: & -\frac{1}{\sigma^2}\tilde{V}^{'}(\beta,\zeta)M\tilde{V}(\beta,\zeta), \end{cases} \tag{4.5}$$

where $W = \text{Blkdiag}(W_1,\ldots,W_R)$, $M = \text{Blkdiag}(M_1,\ldots,M_R)$. Using $e^{\alpha_0 W}Y = Z\beta_0 + C_\lambda\lambda_0 + e^{-\tau_0 M}V$, we can show that $\tilde{V}(\beta_0,\zeta_0) = Q_C(\tau_0)V$. Thus, we can derive $\mathrm{E}\left(S^c(\theta_0)\right)$ as

$$\mathrm{E}\left(S^c(\theta_0)\right) = \begin{cases} \beta: & 0_{k\times 1}, \\ \sigma_\epsilon^2: & -\frac{R}{2\sigma_0^2}, \\ \alpha: & -\mathrm{tr}\left(Q_C(\tau_0)e^{\tau_0 M}We^{-\tau_0 M}\right), \\ \tau: & -\mathrm{tr}\left(Q_C(\tau_0)M\right), \end{cases} \tag{4.6}$$

where $N = n - R$. Using Lemma A.7 in the web appendix, we can determine the order of $\frac{1}{N}\mathrm{E}\left(S^c(\theta_0)\right)$. Thus, we have $\frac{1}{N}\mathrm{tr}\left(Q_C(\tau_0)e^{\tau_0 M}We^{-\tau_0 M}\right) = \frac{1}{N}\mathrm{tr}(e^{\tau_0 M}We^{-\tau_0 M}) - \frac{1}{N}\mathrm{tr}(P_C(\tau_0)e^{\tau_0 M}We^{-\tau_0 M}) = O(1/\min\{n_1,\ldots,n_R\})$ and similarly, $\frac{1}{N}\mathrm{tr}\left(Q_C(\tau_0)M\right) = O(1/\min\{n_1,\ldots,n_R\})$. These order terms indicate that when all group sizes diverge, we can ensure that $\mathrm{plim}_{N\to\infty}\frac{1}{N}S^c(\theta_0) = 0$. However, we may have $\mathrm{plim}_{N\to\infty}\frac{1}{N}S^c(\theta_0) \neq 0$ if some group sizes are not large. To make sure that $\mathrm{plim}_{N\to\infty}\frac{1}{N}S^c(\theta_0) = 0$ holds in all cases, we suggest using the adjusted score functions

$S^{\dagger}(\theta_0) = S^c(\theta_0) - \mathrm{E}\left(S^c(\theta_0)\right)$:

$$S^{\dagger}(\theta) = \begin{cases} \beta: & \frac{1}{\sigma^2} Z' e^{\tau M'} \tilde{V}(\beta,\zeta), \\ \sigma^2: & -\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4} \tilde{V}'(\beta,\zeta)\tilde{V}(\beta,\zeta), \\ \alpha: & -\frac{1}{\sigma^2} Y' e^{\alpha W'} W' e^{\tau M'} \tilde{V}(\beta,\zeta) + \mathrm{tr}\left(Q_C(\tau) e^{\tau M} W e^{-\tau M}\right), \\ \tau: & -\frac{1}{\sigma^2} \tilde{V}'(\beta,\zeta) M \tilde{V}(\beta,\zeta) + \mathrm{tr}\left(Q_C(\tau) M\right). \end{cases} \tag{4.7}$$

For a given ζ value, we can use the adjusted score functions with respect to β and σ_ϵ^2 to get the following estimators:

$$\widehat{\beta}^{\dagger}(\zeta) = \left(Z' e^{\tau M'} Q_C(\tau) e^{\tau M} Z\right)^{-1} Z' e^{\tau M'} Q_C(\tau) e^{\tau M} e^{\alpha W} Y, \tag{4.8}$$

$$\widehat{\sigma}^{\dagger 2}(\zeta) = \frac{1}{N} \widehat{V}'(\zeta) \widehat{V}(\zeta), \tag{4.9}$$

where $\widehat{V}(\zeta) = \tilde{V}(\widehat{\beta}^{\dagger}(\zeta), \zeta) = Q_C(\tau) e^{\tau M} \left(e^{\alpha W} Y - X \widehat{\beta}^{\dagger}(\zeta)\right)$. By substituting $\widehat{\beta}^{\dagger}(\zeta)$ and $\widehat{\sigma}^{\dagger 2}(\zeta)$ into the α and τ elements of $S^{\dagger}(\theta)$, we obtain

$$S^{\dagger c}(\zeta) = \begin{cases} \alpha: & -\frac{1}{\widehat{\sigma}^{\dagger 2}(\zeta)} Y' e^{\alpha W'} W' e^{\tau M'} \widehat{V}(\zeta) + \mathrm{tr}\left(Q_C(\tau) e^{\tau M} W e^{-\tau M}\right), \\ \tau: & -\frac{1}{\widehat{\sigma}^{\dagger 2}(\zeta)} \widehat{V}'(\zeta) M \widehat{V}(\zeta) + \mathrm{tr}\left(Q_C(\tau) M\right). \end{cases} \tag{4.10}$$

Then, we define the M-estimator (ME) $\widehat{\zeta}^{\dagger}$ of ζ_0 by $\widehat{\zeta}^{\dagger} = \mathrm{argsolve}\{S^{\dagger c}(\zeta) = 0\}$. We can use $\widehat{\zeta}^{\dagger}$ to define the MEs $\widehat{\beta}^{\dagger} = \widehat{\beta}^{\dagger}(\widehat{\zeta}^{\dagger})$ and $\widehat{\sigma}^{\dagger 2} = \widehat{\sigma}^{\dagger 2}(\widehat{\zeta}^{\dagger})$. The consistency of $\widehat{\zeta}^{\dagger}$ requires the population counterpart of $S^{\dagger c}(\zeta)$, which we denote by $\overline{S}^{\dagger}(\theta) = \mathrm{E}\left(S^{\dagger}(\theta)\right)$. For a given ζ value, we can derive the following population counterparts of (4.8) and (4.9) from $\overline{S}^{\dagger}(\theta)$ as:

$$\overline{\beta}^{\dagger}(\zeta) = \left(Z' e^{\tau M'} Q_C(\tau) e^{\tau M} Z\right)^{-1} Z' e^{\tau M'} Q_C(\tau) D(\zeta) \mathrm{E}(Y), \tag{4.11}$$

$$\overline{\sigma}^{\dagger 2}(\zeta) = \frac{1}{N} \mathrm{E}\left(\overline{V}'(\zeta) \overline{V}(\zeta)\right), \tag{4.12}$$

where $D(\zeta) = e^{\tau M} e^{\alpha W}$ and $\overline{V}(\zeta) = \tilde{V}(\overline{\beta}^{\dagger}(\zeta), \zeta)$. By substituting $\overline{\beta}^{\dagger}(\zeta)$ and $\overline{\sigma}^{\dagger 2}(\zeta)$ into the α and τ elements of $\overline{S}^{\dagger}(\theta)$, the population counterpart of $S^{\dagger c}(\zeta)$ in (4.10) is given by:

$$\overline{S}^{\dagger c}(\zeta) = \begin{cases} \alpha: & -\frac{1}{\overline{\sigma}^{\dagger 2}(\zeta)} \mathrm{E}\left(Y' e^{\alpha W'} W' e^{\tau M'} \overline{V}(\zeta)\right) + \mathrm{tr}\left(Q_C(\tau) e^{\tau M} W e^{-\tau M}\right), \\ \tau: & -\frac{1}{\overline{\sigma}^{\dagger 2}(\zeta)} \mathrm{E}\left(\overline{V}'(\zeta) M \overline{V}(\zeta)\right) + \mathrm{tr}\left(Q_C(\tau) M\right). \end{cases} \tag{4.13}$$

The consistency of $\widehat{\zeta}^{\dagger}$ requires the following additional assumptions.

Assumption 10. *Z is exogenous, with uniformly bounded elements, and has full column rank.*

Also $\lim_{N\to\infty}\frac{1}{N}Z'e^{\tau M'}Q_C(\tau)e^{\tau M}Z$ *exists and is nonsingular, uniformly in* $\tau\in[-c,c]$.

Assumption 11. $\inf_{\zeta:d(\zeta,\zeta_0)\geq\nu}\left\|\bar{S}^{\dagger c}(\zeta)\right\|>0$ *for every* $\nu>0$*, where* $d(\zeta,\zeta_0)$ *is a measure of distance between* ζ *and* ζ_0.

Assumptions 10 and 11 are the modified versions of Assumptions 5 and 9, respectively. Then, the consistency of $\widehat{\zeta}^{\dagger}$ follows from the uniform convergence $\sup_{\zeta\in\Delta}\frac{1}{N}\left\|S^{\dagger c}(\zeta)-\bar{S}^{\dagger c}(\zeta)\right\|\xrightarrow{p}0$, which is shown in the following theorem.

Theorem 4.1. *Under Assumptions 1, 3, 4, 10 and 11, it follows that* $\widehat{\theta}^{\dagger}\xrightarrow{p}\theta_0$ *as* $N\to\infty$.

Proof. See Section C.1 in the web appendix. □

Next, we will derive the asymptotic distribution of $\widehat{\theta}^{\dagger}$. To that end, we apply the mean value theorem to $S^{\dagger}(\widehat{\theta}^{\dagger})=0$ at θ_0, to obtain $\sqrt{N}(\widehat{\theta}^{\dagger}-\theta_0)=-\left(\frac{1}{N}\frac{\partial S^{\dagger}(\bar{\theta})}{\partial\theta'}\right)^{-1}\frac{1}{\sqrt{N}}S^{\dagger}(\theta_0)$, where $\bar{\theta}$ lies between θ_0 and $\widehat{\theta}^{\dagger}$ elementwise. Let $\phi=Z\beta_0+C_{\lambda}\lambda_0$. By (4.1), Y can be expressed as $Y=e^{-\alpha_0W}(\phi+e^{-\tau_0M}V)$. Then, substituting $Y=e^{-\alpha_0W}(\phi+e^{-\tau_0M}V)$ and $\tilde{V}(\beta_0,\zeta_0)=Q_C(\tau_0)V$ into $S^{\dagger}(\theta_0)$, we can express $S^{\dagger}(\theta_0)$ in terms of linear and quadratic forms of V in the following way,

$$S^{\dagger}(\theta_0)=\begin{cases}\beta: & \frac{1}{\sigma_0^2}Z'(\tau_0)V,\\ \sigma^2: & -\frac{N}{2\sigma_0^2}+\frac{1}{2\sigma_0^4}V'Q_C(\tau_0)V,\\ \alpha: & -\frac{1}{\sigma_0^2}V'T'(\tau_0)Q_C(\tau_0)V-\frac{1}{\sigma_0^2}\phi'e^{\tau_0M'}T'(\tau_0)Q_C(\tau_0)V+\operatorname{tr}\left(Q_C(\tau_0)T(\tau_0)\right),\\ \tau: & -\frac{1}{\sigma_0^2}V'Q_C'(\tau_0)MQ_C(\tau_0)V+\operatorname{tr}\left(Q_C(\tau_0)M\right),\end{cases}\tag{4.14}$$

where $T(\tau_0)=e^{\tau_0M}We^{-\tau_0M}$. The CLT for linear-quadratic forms in Lemma A.4 in the web appendix can be used to establish the asymptotic normality of $\frac{1}{\sqrt{N}}S^{\dagger}(\theta_0)$. Also, our assumptions ensure that $\frac{1}{N}\frac{\partial S^{\dagger}(\bar{\theta})}{\partial\theta'}-\frac{1}{N}\mathrm{E}\left(\frac{\partial S^{\dagger}(\theta_0)}{\partial\theta'}\right)=o_p(1)$. These results lead to the following theorem.

Theorem 4.2. *Under Assumptions 1, 3, 4, 10 and 11, as* $N\to\infty$*, we have*

$$\sqrt{N}(\widehat{\theta}^{\dagger}-\theta_0)\xrightarrow{d}N\left(0,\lim_{N\to\infty}\Psi^{\dagger-1}(\theta_0)\Omega^{\dagger}(\theta_0)\Psi^{\dagger-1'}(\theta_0)\right),\tag{4.15}$$

where $\Psi^{\dagger}(\theta_0)=-\frac{1}{N}\mathrm{E}\left(\frac{\partial S^{\dagger}(\theta_0)}{\partial\theta'}\right)$ *and* $\Omega^{\dagger}(\theta_0)=\mathrm{Var}\left(\frac{1}{\sqrt{N}}S^{\dagger}(\theta_0)\right)$ *are assumed to exist and* $\Psi^{\dagger}(\theta_0)$ *is assumed to be positive definite for sufficiently large* N.

Proof. See Section C.2 in the web appendix. □

An estimator of $\Psi^{\dagger}(\theta_0)$ can be formulated from the observed counterpart given by $H^{\dagger}(\widehat{\theta}^{\dagger}) = -\frac{1}{N}\frac{\partial S^{\dagger}(\theta)}{\partial \theta'}|_{\theta=\widehat{\theta}^{\dagger}}$. Let $H^{\dagger}_{ab}(\theta) = -\frac{1}{N}\frac{\partial S^{\dagger}_a(\theta)}{\partial b'}$ for $a, b \in \{\beta, \sigma^2, \alpha, \tau\}$. Then, we can derive the elements of $H^{\dagger}(\theta)$ as

$$
\begin{aligned}
&NH^{\dagger}_{\beta\beta}(\theta) = \frac{1}{\sigma^2}Z^{'}(\tau)Z(\tau), \quad NH^{\dagger}_{\beta\sigma^2}(\theta) = NH^{\dagger'}_{\sigma^2\beta}(\theta) = \frac{1}{\sigma^4}Z^{'}(\tau)\tilde{V}(\beta,\zeta),\\
&NH^{\dagger}_{\beta\alpha}(\theta) = H^{\dagger'}_{\alpha\beta}(\theta) = -\frac{1}{\sigma^2}Z^{'}(\tau)Y(\zeta), \quad NH^{\dagger}_{\beta\tau}(\theta) = H^{\dagger'}_{\tau\beta}(\theta) = -\frac{1}{\sigma^2}Z^{'}(\tau)M^s\tilde{V}(\beta,\zeta),\\
&NH^{\dagger}_{\sigma^2\sigma^2}(\theta) = -\frac{N}{2\sigma^4} + \frac{1}{\sigma^6}\tilde{V}^{'}(\beta,\zeta)\tilde{V}(\beta,\zeta), \quad NH^{\dagger}_{\sigma^2\alpha}(\theta) = NH^{\dagger}_{\alpha\sigma^2}(\theta) = -\frac{1}{\sigma^4}Y^{'}(\zeta)\tilde{V}(\beta,\zeta),\\
&NH^{\dagger}_{\sigma^2\tau}(\theta) = NH^{\dagger}_{\tau\sigma^2}(\theta) = -\frac{1}{\sigma^4}\tilde{V}^{'}(\beta,\zeta)M\tilde{V}(\beta,\zeta), \quad NH^{\dagger}_{\alpha\alpha}(\theta) = \frac{1}{\sigma^2}\left(Y_2^{'}(\zeta)\tilde{V}(\beta,\zeta) + Y^{'}(\zeta)Y(\zeta)\right),\\
&NH^{\dagger}_{\alpha\tau}(\theta) = \frac{1}{\sigma^2}Y^{'}(\zeta)M^s\tilde{V}(\beta,\zeta) + \operatorname{tr}\left(P_C(\tau)M^sQ_C(\tau)T(\tau)\right),\\
&NH^{\dagger}_{\tau\alpha}(\theta) = \frac{1}{\sigma^2}Y^{'}(\zeta)M^s\tilde{V}(\beta,\zeta),\\
&NH^{\dagger}_{\tau\tau}(\theta) = \frac{1}{\sigma^2}\tilde{V}^{'}(\beta,\zeta)B^{'}(\tau)M^s\tilde{V}(\beta,\zeta) + \operatorname{tr}\left(Q_C(\tau)MP_C(\tau)M^s\right),
\end{aligned}
$$

where $Z(\tau) = Q_C(\tau)e^{\tau M}Z$, $Y(\zeta) = Q_C(\tau)e^{\tau M}We^{\alpha W}Y$, $Y_2(\zeta) = Q_C(\tau)e^{\tau M}W^2e^{\alpha W}Y$, $B(\tau) = Q_C(\tau)M - P_C(\tau)M^{'}$, and $T(\tau) = e^{\tau M}We^{-\tau M}$. In the proof of Theorem 4.2, we show that $H^{\dagger}(\widehat{\theta}^{\dagger}) = \Psi^{\dagger}(\theta_0) + o_p(1)$.

Define $\Omega^{\dagger}_{ab}(\theta_0) = \frac{1}{N}\mathrm{E}\left(S^{\dagger}_a(\theta_0)S^{\dagger'}_b(\theta_0)\right)$ for $a, b \in \{\beta, \sigma^2, \alpha, \tau\}$. Let $q_1 = \mathrm{vec}_D\left(Q_C(\tau_0)\right)$, $\mathcal{Q}_2(\tau_0) = Q_C(\tau_0)T(\tau_0)$, $\mathcal{Q}_3(\tau_0) = Q_C(\tau_0)e^{\tau_0 M}W$, $\mathcal{Q}_4(\tau_0) = Q_C(\tau_0)MQ_C(\tau_0)$ and $q_i = \mathrm{vec}_D\left(\mathcal{Q}_i(\tau_0)\right)$ for $i = 2, 3, 4$. Let $\rho = \mathrm{E}(v_i^3)/\sigma_0^3$ and $\kappa = \mathrm{E}(v_i^4)/\sigma_0^4 - 3$ be the skewness and excess kurtosis parameters, respectively. Then, by Lemma A.2 in the web appendix, the closed-form expressions for the elements of $\Omega^{\dagger}(\theta_0)$ can be derived as:

$$
\begin{aligned}
&N\Omega^{\dagger}_{\beta\beta}(\theta_0) = \frac{1}{\sigma_0^2} Z'(\tau_0) Z(\tau_0), \quad N\Omega^{\dagger}_{\beta\sigma^2}(\theta_0) = \frac{\rho}{2\sigma_0^3} Z'(\tau_0) q_1, \\
&N\Omega^{\dagger}_{\beta\alpha}(\theta_0) = -\frac{1}{\sigma_0^2} Z'(\tau_0) \mathcal{Q}_3(\tau_0)\phi - \frac{\rho}{\sigma_0} Z'(\tau_0) q_2, \\
&N\Omega^{\dagger}_{\beta\tau}(\theta_0) = -\frac{\rho}{\sigma_0} Z'(\tau_0) q_4, \quad N\Omega^{\dagger}_{\sigma^2\sigma^2}(\theta_0) = \frac{1}{4\sigma_0^4}(\kappa q_1' q_1 + 2N), \\
&N\Omega^{\dagger}_{\sigma^2\alpha}(\theta_0) = -\frac{\rho}{2\sigma_0^3} q_1' \mathcal{Q}_3(\tau_0)\phi - \frac{1}{2\sigma_0^2}\left(\kappa q_1' q_2 + 2\mathrm{tr}(\mathcal{Q}_2(\tau_0))\right), \\
&N\Omega^{\dagger}_{\sigma^2\tau}(\theta_0) = -\frac{1}{2\sigma_0^2}\left(\kappa q_1' q_4 + 2\mathrm{tr}(Q_C(\tau_0)M)\right), \\
&N\Omega^{\dagger}_{\alpha\alpha}(\theta_0) = \frac{1}{\sigma_0^2}\phi' \mathcal{Q}_3'(\tau_0)\mathcal{Q}_3(\tau_0)\phi + \frac{2\rho}{\sigma_0} q_2' \mathcal{Q}_3(\tau_0)\phi + \kappa q_2' q_2 + \mathrm{tr}\left(\mathcal{Q}_2(\tau_0)\mathcal{Q}_2^s(\tau_0)\right), \\
&N\Omega^{\dagger}_{\alpha\tau}(\theta_0) = \frac{\rho}{\sigma_0} q_4' \mathcal{Q}_3(\tau_0)\phi + \kappa q_2' q_4 + \mathrm{tr}\left(\mathcal{Q}_2(\tau_0)\mathcal{Q}_4^s(\tau_0)\right), \\
&N\Omega^{\dagger}_{\tau\tau}(\theta_0) = \kappa q_4' q_4 + \mathrm{tr}\left(\mathcal{Q}_4(\tau_0)\mathcal{Q}_4^s(\tau_0)\right).
\end{aligned}
\tag{4.16}
$$

Let $\widehat{\lambda}^{\dagger} = \widehat{\lambda}(\widehat{\beta}^{\dagger}, \widehat{\zeta}^{\dagger})$ be the plug-in estimator of λ_0 obtained from (4.3), and $\widehat{\rho}$ and $\widehat{\kappa}$ be consistent estimators of ρ and κ, respectively. The plug-in estimator of $\Omega^{\dagger}(\theta_0)$ is then given by $\Omega^{\dagger}(\widehat{\theta}^{\dagger}) = \Omega^{\dagger}(\theta)\big|_{\theta=\widehat{\theta}^{\dagger}, \lambda=\widehat{\lambda}^{\dagger}, \rho=\widehat{\rho}, \kappa=\widehat{\kappa}}$. The next theorem shows that this plug-in estimator is a biased estimator of $\Omega^{\dagger}(\theta_0)$ under our stated assumptions.

Theorem 4.3. *Under Assumptions 1, 3, 4, 10 and 11, as $N \to \infty$, we have*

$$
\Omega^{\dagger}(\widehat{\theta}^{\dagger}) = \Omega^{\dagger}(\theta_0) + \mathrm{Bias}^{\dagger}(\tau_0) + o_p(1), \tag{4.17}
$$

where $\mathrm{Bias}^{\dagger}(\tau_0)$ *is an* $(k+3)\times(k+3)$ *matrix with zero entries everywhere except the* (α,α) *entry, which is* $\frac{1}{N}\mathrm{tr}\left(P_C(\tau_0)\mathcal{Q}_2'(\tau_0)\mathcal{Q}_2(\tau_0)\right)$.

Proof. See Section C.3 in the web appendix. □

The bias term arises since we may not be able to consistently estimate the group fixed effects λ_0. Under our assumptions, both $\mathcal{Q}_2(\tau_0)$ and $P_C(\tau_0)$ are bounded in row sum and column sum matrix norms. Using Lemma A.7 in the web appendix, we can show that $\frac{1}{N}\mathrm{tr}\left(P_C(\tau_0)\mathcal{Q}_2'(\tau_0)\mathcal{Q}_2(\tau_0)\right) = O(1/\min\{n_1,\ldots,n_R\})$. Thus, when all group sizes are large, the bias term will become negligible. However, in settings with some fixed group sizes, the bias correction is necessary for valid inference. Therefore, we suggest using the bias corrected estimator $\widehat{\Omega}^{\dagger} = \Omega^{\dagger}(\widehat{\theta}^{\dagger}) - \mathrm{Bias}^{\dagger}(\widehat{\tau}^{\dagger})$ to have valid inference in all cases.

Note that the plug-in estimator $\Omega^{\dagger}(\widehat{\theta}^{\dagger})$ still requires the consistent estimators of ρ and κ. Recall that $\tilde{V}(\beta_0,\zeta_0) = Q_C(\tau_0)e^{\tau_0 M}(e^{\alpha_0 W}Y - Z\beta_0) = Q_C(\tau_0)V$. Then, from $\tilde{V}(\beta_0,\zeta_0) =$

$Q_C(\tau_0)V$, we have $\tilde{v}_j = q_{j1}v_1 + \ldots + q_{jn}v_n$, where q_{jh} is the (j,h)th element of $Q_C(\tau_0)$, and v_j and $\tilde{v}_j$ are the jth element of V and $\tilde{V}(\beta_0, \zeta_0)$, respectively. Since v_j's are $i.i.d$, we have $\mathrm{E}(\tilde{v}_j^3) = \sum_{h=1}^{N} q_{jh}^3 \mathrm{E}(v_h^3) = \sigma_0^3 \rho \sum_{h=1}^{n} q_{jh}^3$. Summing $\mathrm{E}(\tilde{v}_j^3)$ over j and solving for ρ, we obtain

$$\rho = \frac{\sum_{j=1}^{n} \mathrm{E}(\tilde{v}_j^3)}{\sigma_0^3 \sum_{j=1}^{n} \sum_{h=1}^{n} q_{jh}^3}. \tag{4.18}$$

This result suggests the following sample counterpart as a consistent estimator for ρ:

$$\widehat{\rho} = \frac{\sum_{j=1}^{n} \widehat{v}_j^3}{\widehat{\sigma}^{\dagger 3} \sum_{j=1}^{n} \sum_{h=1}^{n} \widehat{q}_{jh}^3}, \tag{4.19}$$

where $\widehat{v}_j$ is the jth element of $\widehat{V}(\widehat{\beta}^{\dagger}, \widehat{\zeta}^{\dagger}) = Q_C(\widehat{\tau}^{\dagger}) e^{\widehat{\tau}^{\dagger} M}(e^{\widehat{\alpha}^{\dagger} W} Y - Z\widehat{\beta}^{\dagger})$ and $\widehat{q}_{jh}$ is the (j,h)th element of $Q_C(\widehat{\tau}^{\dagger})$. Similarly, from $\tilde{v}_j = q_{j1}v_1 + \ldots + q_{jn}v_n$, we obtain

$$\begin{aligned}
\mathrm{E}(\tilde{v}_j^4) &= \sum_{h=1}^{n} q_{jh}^4 \mathrm{E}(v_h^4) + 3\sigma_0^4 \sum_{h=1}^{n} \sum_{l=1}^{n} q_{jh}^2 q_{jl}^2 - 3\sigma_0^4 \sum_{h=1}^{n} q_{jh}^4 \\
&= \kappa \sum_{h=1}^{n} q_{jh}^4 \sigma_0^4 + 3\sigma_0^4 \sum_{h=1}^{n} \sum_{l=1}^{n} q_{jh}^2 q_{jl}^2.
\end{aligned} \tag{4.20}$$

Summing $\mathrm{E}(\tilde{v}_j^4)$ over j and solving for κ, we obtain

$$\kappa = \frac{\sum_{j=1}^{n} \mathrm{E}(\tilde{v}_j^4) - 3\sigma_0^4 \sum_{j=1}^{n} \sum_{h=1}^{n} \sum_{l=1}^{n} q_{jh}^2 q_{jl}^2}{\sigma_0^4 \sum_{j=1}^{n} \sum_{h=1}^{n} q_{jh}^4}. \tag{4.21}$$

Thus, we can consider the following sample counterpart of this equation as a consistent estimator for κ:

$$\widehat{\kappa} = \frac{\sum_{j=1}^{n} \widehat{v}_j^4 - 3\widehat{\sigma}^{\dagger 4} \sum_{j=1}^{n} \sum_{h=1}^{n} \sum_{l=1}^{n} \widehat{q}_{jh}^2 \widehat{q}_{jl}^2}{\widehat{\sigma}^{\dagger 4} \sum_{j=1}^{n} \sum_{h=1}^{n} \widehat{q}_{jh}^4}. \tag{4.22}$$

4.2 Estimation under heteroskedasticity

Recall that $D(\zeta_0) = e^{\tau_0 M} e^{\alpha_0 W}$, $D(\zeta_0)Y = e^{\tau_0 M}\phi + V$, $T(\tau_0) = e^{\tau_0 M} W e^{-\tau_0 M}$ and $\tilde{V}(\beta_0, \zeta_0) = Q_C(\tau_0)V$. Let $\tilde{V} = \tilde{V}(\beta_0, \zeta_0)$ and $\overline{T}'(\tau_0) = \mathrm{Diag}\left(T'(\tau_0) Q_C(\tau_0)\right) \mathrm{Diag}\left(Q_C(\tau_0)\right)^{-1}$. Also, note that $\mathrm{tr}(AB) = \mathrm{tr}(A\,\mathrm{Diag}(B))$ for a diagonal matrix A and any conformable matrix B. We will use this fact to adjust the score functions. Under Assumption 8, the α element of

$\mathrm{E}\left(S^c(\theta_0)\right)$ can be determined as

$$\begin{aligned}
&\mathrm{E}\left(Y^{'}e^{\alpha_0 W^{'}}W^{'}e^{\tau_0 M^{'}}\tilde{V}\right) = \mathrm{E}\left(Y^{'}D^{'}(\zeta_0)T^{'}(\tau_0)Q_C(\tau_0)V\right) = \mathrm{tr}\left(\Sigma T^{'}(\tau_0)Q_C(\tau_0)\right)\\
&= \mathrm{tr}\left(\Sigma\,\mathrm{Diag}\left(T^{'}(\tau_0)Q_C(\tau_0)\right)\right) = \mathrm{tr}\left(\Sigma\,\mathrm{Diag}\left(T^{'}(\tau_0)Q_C(\tau_0)\right)\mathrm{Diag}\left(Q_C(\tau_0)\right)^{-1}Q_C(\tau_0)\right)\\
&= \mathrm{tr}\left(\Sigma\overline{T}^{'}(\tau_0)Q_C(\tau_0)\right) = \mathrm{E}\left(Y^{'}D^{'}(\zeta_0)\overline{T}^{'}(\tau_0)Q_C(\tau_0)V\right). \qquad (4.23)
\end{aligned}$$

In (4.23), subtracting the last element from the second element yields

$$\begin{aligned}
&\mathrm{E}\left(Y^{'}D^{'}(\zeta_0)T^{'}(\tau_0)Q_C(\tau_0)V\right) - \mathrm{E}\left(Y^{'}D^{'}(\zeta_0)\overline{T}^{'}(\tau_0)Q_C(\tau_0)V\right) = 0,\\
&\Longrightarrow\ \mathrm{E}\left(Y^{'}D^{'}(\zeta_0)\left(T^{'}(\tau_0) - \overline{T}^{'}(\tau_0)\right)Q_C(\tau_0)V\right) = 0. \qquad (4.24)
\end{aligned}$$

We suggest using the sample counterpart of (4.24) as the adjusted robust score function for the α element:

$$-Y^{'}D^{'}(\zeta)\left(T^{'}(\tau) - \overline{T}^{'}(\tau)\right)\tilde{V}(\beta,\zeta). \qquad (4.25)$$

Let $\overline{\mathcal{Q}}_4(\tau_0) = \mathrm{Diag}\left(\mathcal{Q}_4(\tau_0)\right)\mathrm{Diag}\left(Q_C(\tau_0)\right)^{-1}$, where $\mathcal{Q}_4(\tau_0) = Q_C(\tau_0)MQ_C(\tau_0)$. Then, applying similar steps to the τ element of $\mathrm{E}\left(S^c(\theta_0)\right)$ yields

$$\begin{aligned}
&\mathrm{E}\left(\tilde{V}^{'}M\tilde{V}\right) = \mathrm{E}\left(V^{'}Q_C(\tau_0)MQ_C(\tau_0)V\right) = \mathrm{tr}\left(\Sigma\mathcal{Q}_4(\tau_0)\right) = \mathrm{tr}\left(\Sigma\,\mathrm{Diag}(\mathcal{Q}_4(\tau_0))\right)\\
&= \mathrm{tr}\left(\Sigma\,\mathrm{Diag}\left(\mathcal{Q}_4(\tau_0)\right)\mathrm{Diag}\left(Q_C(\tau_0)\right)^{-1}Q_C(\tau_0)\right) = \mathrm{E}\left(V^{'}\overline{\mathcal{Q}}_4(\tau_0)Q_C(\tau_0)V\right). \qquad (4.26)
\end{aligned}$$

Subtracting the last element from the second element in (4.26) yields

$$\begin{aligned}
&\mathrm{E}\left(V^{'}Q_C(\tau_0)MQ_C(\tau_0)V\right) - \mathrm{E}\left(V^{'}\overline{\mathcal{Q}}_4(\tau_0)Q_C(\tau_0)V\right) = 0,\\
&\Longrightarrow\ \mathrm{E}\left(V^{'}\left(\mathcal{Q}_4(\tau_0) - \overline{\mathcal{Q}}_4(\tau_0)\right)Q_C(\tau_0)V\right) = 0,\\
&\Longrightarrow\ \mathrm{E}\left((e^{\alpha_0 W}Y - Z\beta_0)^{'}e^{\tau_0 M^{'}}\left(\mathcal{Q}_4(\tau_0) - \overline{\mathcal{Q}}_4(\tau_0)\right)Q_C(\tau_0)V\right) = 0 \qquad (4.27)
\end{aligned}$$

Based on (4.27), we suggest using the following adjusted robust score function with respect to τ:

$$-(e^{\alpha W}Y - Z\beta)^{'}e^{\tau M^{'}}\left(\mathcal{Q}_4(\tau) - \overline{\mathcal{Q}}_4(\tau)\right)\tilde{V}(\beta,\zeta). \qquad (4.28)$$

Combining (4.25) and (4.28), we suggest the following robust versions of the score functions for consistent estimation under Assumption 8:

$$S^{\ddagger}(\beta,\zeta)=\begin{cases}\beta: & Z^{'}(\tau)\tilde{V}(\beta,\zeta),\\ \alpha: & -Y^{'}D^{'}(\zeta)\left(T^{'}(\tau)-\overline{T}^{'}(\tau)\right)\tilde{V}(\beta,\zeta),\\ \tau: & -\left(e^{\alpha W}Y-Z\beta\right)^{'}e^{\tau M^{'}}\left(\mathcal{Q}_4(\tau)-\overline{\mathcal{Q}}_4(\tau)\right)\tilde{V}(\beta,\zeta).\end{cases}\tag{4.29}$$

Note that from (4.23), we have $\frac{1}{N}\mathrm{E}\left(Y^{'}e^{\alpha_0 W^{'}}W^{'}e^{\tau_0 M^{'}}\tilde{V}\right)=\frac{1}{N}\mathrm{tr}\left(\Sigma T^{'}(\tau_0)Q_C(\tau_0)\right)=\frac{1}{N}\mathrm{tr}\left(\Sigma T^{'}(\tau_0)\right)-\frac{1}{N}\mathrm{tr}\left(\Sigma T^{'}(\tau_0)P_C(\tau_0)\right)=O(1)+O(1/\min\{n_1,\ldots,n_R\})=O(1)$ by Lemma A.7 in the web appendix. If M and W are commutative, then $\mathrm{tr}\left(\Sigma T^{'}(\tau_0)\right)=0$, which yields $\frac{1}{N}\mathrm{E}\left(Y^{'}e^{\alpha_0 W^{'}}W^{'}e^{\tau_0 M^{'}}\tilde{V}\right)=O(1/\min\{n_1,\ldots,n_R\})$. Also, from (4.26), we have the following: $\frac{1}{N}\mathrm{E}\left(\tilde{V}^{'}M\tilde{V}\right)=\frac{1}{N}\mathrm{tr}\left(\Sigma\mathcal{Q}_4(\tau_0)\right)=\frac{1}{N}\mathrm{tr}\left(\Sigma P_C(\tau_0)MP_C(\tau_0)\right)-\frac{1}{N}\mathrm{tr}\left(\Sigma MP_C(\tau_0)\right)-\frac{1}{N}\mathrm{tr}\left(\Sigma P_C(\tau_0)M\right)=O(1/\min\{n_1,\ldots,n_R\})$ by Lemma A.7 in the web appendix. These results indicate that the necessary condition $\mathrm{plim}_{N\to\infty}\frac{1}{N}S^c(\theta_0)=0$ can only holds if W and M are commutative and all group sizes are large. To ensure that $\mathrm{plim}_{N\to\infty}\frac{1}{N}S^c(\theta_0)=0$ holds in all cases, we suggest using $S^{\ddagger}(\beta,\zeta)$ for estimation.

To derive the robust M-estimator from (4.29), we can similarly first solve for β for a given ζ from the score function with respect to β, which is the same as (4.8):

$$\widehat{\beta}^{\ddagger}(\zeta)=\widehat{\beta}^{\dagger}(\zeta)=\left(Z^{'}e^{\tau M^{'}}Q_C(\tau)e^{\tau M}Z\right)^{-1}Z^{'}e^{\tau M^{'}}Q_C(\tau)e^{\tau M}e^{\alpha W}Y.\tag{4.30}$$

Then, substituting $\widehat{\beta}^{\ddagger}(\zeta)$ into the α and τ elements of $S^{\ddagger}(\beta,\zeta)$, we obtain the concentrated robust score functions:

$$S^{\ddagger c}(\zeta)=\begin{cases}\alpha: & -Y^{'}D^{'}(\zeta)\left(T^{'}(\tau)-\overline{T}^{'}(\tau)\right)\widehat{V}(\zeta),\\ \tau: & -\left(e^{\alpha W}Y-Z\widehat{\beta}^{\ddagger}(\zeta)\right)^{'}e^{\tau M^{'}}\left(\mathcal{Q}_4(\tau)-\overline{\mathcal{Q}}_4(\tau)\right)\widehat{V}(\zeta),\end{cases}\tag{4.31}$$

where $\widehat{V}(\zeta)=\tilde{V}(\widehat{\beta}^{\ddagger}(\zeta),\zeta)$. Then, we define the RME $\widehat{\zeta}^{\ddagger}$ of ζ_0 by $\widehat{\zeta}^{\ddagger}=\mathrm{argsolve}\{S^{\ddagger c}(\zeta)=0\}$. From (4.30), we can define the RME $\widehat{\beta}^{\ddagger}=\widehat{\beta}^{\ddagger}(\widehat{\zeta}^{\ddagger})$ of β_0.

Let $\overline{S}^{\ddagger}(\beta,\zeta)=\mathrm{E}\left(S^{\ddagger}(\beta,\zeta)\right)$ be the population counterpart of the robust score functions. For a given ζ value, we can derive the estimator $\overline{\beta}^{\ddagger}(\zeta)=\overline{\beta}^{*}(\zeta)$ in (4.11), which can be substituted into the α and τ elements of $\overline{S}^{\ddagger}(\beta,\zeta)$ to obtain:

$$\overline{S}^{\ddagger c}(\zeta)=\begin{cases}\alpha: & -\mathrm{E}\left(Y'D'(\zeta)\left(T'(\tau)-\overline{T}'(\tau)\right)\overline{V}(\zeta)\right),\\ \tau: & -\mathrm{E}\left(\left(e^{\alpha W}Y-Z\overline{\beta}^{\ddagger}(\zeta)\right)'e^{\tau M'}\left(\mathcal{Q}_4(\tau)-\overline{\mathcal{Q}}_4(\tau)\right)\overline{V}(\zeta)\right),\end{cases} \tag{4.32}$$

where $\overline{V}(\zeta)=\tilde{V}(\overline{\beta}^{\ddagger}(\zeta),\zeta)$. Let $\omega=(\beta',\zeta')'$ and $\widehat{\omega}^{\ddagger}=(\widehat{\beta}^{\ddagger'},\widehat{\zeta}^{\ddagger'})'$. To show that $\widehat{\omega}^{\ddagger}$ is a consistent estimator of ω_0, we need the following identification assumption.

Assumption 12. $\inf_{\zeta:\,d(\zeta,\zeta_0)\geq\vartheta}\left\|\overline{S}^{\ddagger c}(\zeta)\right\|>0$ *for every* $\vartheta>0$*, where* $d(\zeta,\zeta_0)$ *is a measure of distance between* ζ *and* ζ_0*.*

The following theorem shows that $\widehat{\omega}^{\ddagger}$ is a consistent estimator of ω_0.

Theorem 4.4. *Under Assumptions 3, 4, 8, 10 and 12, as* $N\to\infty$*, we have* $\widehat{\omega}^{\ddagger}\xrightarrow{p}\omega_0$*.*

Proof. See Section C.4 in the web appendix. □

Recall that $\tilde{V}(\beta_0,\zeta_0)=Q_C(\tau_0)V$ and $Y=e^{-\alpha_0 W}\left(\phi+e^{-\tau_0 M}V\right)$. By substituting these two terms into $S^{\ddagger}(\omega_0)$, we have can express $S^{\ddagger}(\omega_0)$ in terms of linear and quadratic forms of V:

$$S^{\ddagger}(\omega_0)=\begin{cases}\beta: Z'(\tau_0)V,\\ \alpha: -\phi'e^{\tau_0 M'}\left(T'(\tau_0)-\overline{T}'(\tau_0)\right)Q_C(\tau_0)V-V'\left(T'(\tau_0)-\overline{T}'(\tau_0)\right)Q_C(\tau_0)V,\\ \tau: -\lambda_0'C'(\tau_0)\left(\mathcal{Q}_4(\tau_0)-\overline{\mathcal{Q}}_4(\tau_0)\right)Q_C(\tau_0)V-V'\left(\mathcal{Q}_4(\tau_0)-\overline{\mathcal{Q}}_4(\tau_0)\right)Q_C(\tau_0)V.\end{cases} \tag{4.33}$$

Then, the asymptotic normality of $\frac{1}{\sqrt{N}}S^{\ddagger}(\omega_0)$ follows from the CLT in Lemma A.4 in the web appendix. The next theorem shows the asymptotic distribution of $\widehat{\omega}^{\ddagger}$.

Theorem 4.5. *Under Assumptions 3, 4, 8, 10 and 12, as* $n\to\infty$*,*

$$\sqrt{N}\left(\widehat{\omega}^{\ddagger}-\omega_0\right)\xrightarrow{d}N\left(0,\lim_{N\to\infty}\Psi^{\ddagger-1}(\omega_0)\Omega^{\ddagger}(\omega_0)\Psi^{\ddagger-1'}(\omega_0)\right), \tag{4.34}$$

where $\Psi^{\ddagger}(\omega_0)=-\frac{1}{N}\mathrm{E}\left(\frac{\partial S^{\ddagger}(\omega_0)}{\partial\omega'}\right)$ *and* $\Omega^{\ddagger}(\omega_0)=\mathrm{Var}\left(\frac{1}{\sqrt{N}}S^{\ddagger}(\omega_0)\right)$ *are assumed to exist and* $\Psi^{\ddagger}(\omega_0)$ *is assumed to be positive definite for sufficiently large* N*.*

Proof. See Section C.5 in the web appendix. □

We can use $H^{\ddagger}(\widehat{\omega}^{\ddagger}) = -\frac{1}{N}\frac{\partial S^{\ddagger}(\omega)}{\partial \omega'}|_{\omega=\widehat{\omega}^{\ddagger}}$ to estimate $\Psi^{\ddagger}(\omega_0)$. Let $H^{\ddagger}_{ab}(\omega) = -\frac{1}{N}\frac{\partial S^{\ddagger}_a(\omega)}{\partial b'}$ for $a, b \in \{\beta, \alpha, \tau\}$. Then, we determine the elements of $H^{\ddagger}(\omega)$ in the following.

$$
\begin{aligned}
&NH^{\ddagger}_{\beta\beta}(\omega) = Z'(\tau)Z(\tau), \quad NH^{\ddagger}_{\beta\alpha}(\omega) = -Z'(\tau)Y(\zeta), \quad NH^{\ddagger}_{\beta\tau}(\omega) = -Z'(\tau)M^s\tilde{V}(\beta,\zeta),\\
&NH^{\ddagger}_{\alpha\beta}(\omega) = -Y'D'(\zeta)\left(T'(\tau) - \overline{T}'(\tau)\right)Z(\tau),\\
&NH^{\ddagger}_{\alpha\alpha}(\omega) = Y'W'D'(\zeta)\left(T'(\tau) - \overline{T}'(\tau)\right)\tilde{V}(\beta,\zeta) + Y'D'(\zeta)\left(T'(\tau) - \overline{T}'(\tau)\right)Y(\zeta),\\
&NH^{\ddagger}_{\alpha\tau}(\omega) = Y'D'(\zeta)\left(M'\left(T'(\tau) - \overline{T}'(\tau)\right) + \dot{T}'(\tau) - \dot{\overline{T}}'(\tau) + \left(T'(\tau) - \overline{T}'(\tau)\right)B(\tau)\right)\tilde{V}(\beta,\zeta),\\
&NH^{\ddagger}_{\tau\beta}(\omega) = -\tilde{V}'(\beta,\zeta)\left(\mathcal{Q}_4'(\tau) - \overline{\mathcal{Q}}_4'(\tau)\right)e^{\tau M}Z - \left(e^{\alpha W}Y - Z\beta\right)'e^{\tau M'}\left(\mathcal{Q}_4(\tau) - \overline{\mathcal{Q}}_4(\tau)\right)Z(\tau),\\
&NH^{\ddagger}_{\tau\alpha}(\omega) = Y'W'D'(\zeta)\left(\mathcal{Q}_4(\tau) - \overline{\mathcal{Q}}_4(\tau)\right)\tilde{V}(\beta,\zeta) + \left(e^{\alpha W}Y - Z\beta\right)'e^{\tau M'}\left(\mathcal{Q}_4(\tau) - \overline{\mathcal{Q}}_4(\tau)\right)Y(\zeta),\\
&NH^{\ddagger}_{\tau\tau}(\omega) = \left(e^{\alpha W}Y - Z\beta\right)'e^{\tau M'}\left(M'\left(\mathcal{Q}_4(\tau) - \overline{\mathcal{Q}}_4(\tau)\right) + \dot{\mathcal{Q}}_4(\tau) - \dot{\overline{\mathcal{Q}}}_4(\tau) + \left(\mathcal{Q}_4(\tau) - \overline{\mathcal{Q}}_4(\tau)\right)B(\tau)\right)\tilde{V},
\end{aligned}
$$

where $\dot{\overline{T}}'(\tau) = \mathrm{Diag}\left(\dot{T}'(\tau)Q_C(\tau) + T'(\tau)\dot{Q}_C(\tau)\right)\mathrm{Diag}\left(Q_C(\tau)\right)^{-1} - \overline{T}(\tau)\,\mathrm{Diag}\left(\dot{Q}_C(\tau)\right)\mathrm{Diag}\left(Q_C(\tau)\right)^{-1}$, $\dot{T}'(\tau) = T'(\tau)M' - M'T'(\tau)$, $\dot{\overline{\mathcal{Q}}}_4(\tau) = \mathrm{Diag}\left(\dot{\mathcal{Q}}_4(\tau)\right)\mathrm{Diag}\left(Q_C(\tau)\right)^{-1} - \overline{\mathcal{Q}}_4(\tau)\,\mathrm{Diag}\left(\dot{Q}_C(\tau)\right)\mathrm{Diag}\left(Q_C(\tau)\right)^{-1}$, $\dot{\mathcal{Q}}_4(\tau) = \dot{Q}_C(\tau)MQ_C(\tau) + Q_C(\tau)M\dot{Q}_C(\tau)$ and $\dot{Q}_C(\tau) = -\left(Q_C(\tau)MP_C(\tau) + P_C(\tau)M'Q_C(\tau)\right)$. As shown in the proof of Theorem 4.5, $H^{\ddagger}(\widehat{\omega}^{\ddagger})$ is a consistent estimator of $\Psi^{\ddagger}(\omega_0)$.

For $\Omega^{\ddagger}(\omega_0)$, we first derive its closed-form expression using Lemma A.2 in the web appendix:

$$
\begin{aligned}
N\Omega^{\ddagger}_{\beta\omega}(\omega_0) &= N\left(\Omega^{\ddagger}_{\beta\beta}(\omega_0), \quad \Omega^{\ddagger}_{\beta\alpha}(\omega_0), \quad \Omega^{\ddagger}_{\beta\tau}(\omega_0)\right)\\
&= \left(Z'(\tau_0)\Sigma Z(\tau_0), \quad -Z'(\tau_0)\Sigma\overline{\overline{T}}(\tau_0)e^{\tau_0 M}\phi, \quad -Z'(\tau_0)\Sigma\overline{\overline{\mathcal{Q}}}(\tau_0)C(\tau_0)\lambda_0\right),\\
N\Omega^{\ddagger}_{\alpha\alpha}(\omega_0) &= \phi'e^{\tau_0 M'}\overline{\overline{T}}'(\tau_0)\Sigma\overline{\overline{T}}(\tau_0)e^{\tau_0 M}\phi + \mathrm{tr}\left(\Sigma\overline{\overline{T}}(\tau_0)\Sigma\overline{\overline{T}}^s(\tau_0)\right),\\
N\Omega^{\ddagger}_{\alpha\tau}(\omega_0) &= \phi'e^{\tau_0 M'}\overline{\overline{T}}'(\tau_0)\Sigma\overline{\overline{\mathcal{Q}}}(\tau_0)C(\tau_0)\lambda_0 + \mathrm{tr}\left(\Sigma\overline{\overline{T}}(\tau_0)\Sigma\overline{\overline{\mathcal{Q}}}^s(\tau_0)\right),\\
N\Omega^{\ddagger}_{\tau\tau}(\omega_0) &= \lambda_0'C'(\tau_0)\overline{\overline{\mathcal{Q}}}'(\tau_0)\Sigma\overline{\overline{\mathcal{Q}}}(\tau_0)C(\tau_0)\lambda_0 + \mathrm{tr}\left(\Sigma\overline{\overline{\mathcal{Q}}}(\tau_0)\Sigma\overline{\overline{\mathcal{Q}}}^s(\tau_0)\right),
\end{aligned}
$$

where $\overline{\overline{T}}(\tau_0) = Q_C(\tau_0)\left(T(\tau_0) - \overline{T}(\tau_0)\right)$ and $\overline{\overline{\mathcal{Q}}}(\tau_0) = Q_C(\tau_0)\left(\mathcal{Q}_4'(\tau_0) - \overline{\mathcal{Q}}_4'(\tau_0)\right)$. For convenience of exposition, let us write $\Omega^{\ddagger}(\omega_0)$ as $\Omega^{\ddagger}(\omega_0, \lambda_0, \Sigma)$. Let $\widehat{\lambda}(\widehat{\beta}^{\ddagger}, \widehat{\zeta}^{\ddagger})$, and $\Omega^{\ddagger}(\widehat{\omega}^{\ddagger}, \widehat{\lambda}^{\ddagger}, \Sigma)$ be the plug-in estimator of $\Omega^{\ddagger}(\omega_0)$ given Σ. This plug-in estimator has bias as shown in the following following theorem.

Theorem 4.6. *Under Assumptions 3, 4, 8, 10 and 12, we have*

$$
\Omega^{\ddagger}(\widehat{\omega}^{\ddagger}, \widehat{\lambda}^{\ddagger}, \Sigma) = \Omega^{\ddagger}(\omega_0, \lambda_0, \Sigma) + \mathrm{Bias}^{\ddagger}_{\lambda}(\tau_0, \Sigma) + o_p(1),
$$

where $\text{Bias}^{\ddagger}_{\lambda}(\tau_0,\Sigma)$ *is an* $(k+2)\times(k+2)$ *matrix given as*

$$\text{Bias}^{\ddagger}_{\lambda}(\tau_0,\Sigma) = \begin{pmatrix} 0_{k\times k} & 0_{k\times 1} & 0_{k\times 1} \\ 0_{1\times k} & \text{Bias}^{\ddagger}_{\lambda,\alpha\alpha}(\tau_0,\Sigma) & \text{Bias}^{\ddagger}_{\lambda,\alpha\tau}(\tau_0,\Sigma) \\ 0_{1\times k} & \text{Bias}^{\ddagger}_{\lambda,\tau\alpha}(\tau_0,\Sigma) & \text{Bias}^{\ddagger}_{\lambda,\tau\tau}(\tau_0,\Sigma) \end{pmatrix},$$

with

$$\text{Bias}^{\ddagger}_{\lambda,\alpha\alpha}(\tau_0,\Sigma) = \frac{1}{N}\text{tr}\left(\Sigma P_C(\tau_0)\overline{\overline{T}}'(\tau_0)\Sigma\overline{\overline{T}}(\tau_0)P_C(\tau_0)\right),$$
$$\text{Bias}^{\ddagger}_{\lambda,\alpha\tau}(\tau_0,\Sigma) = \text{Bias}^{\ddagger}_{\lambda,\tau\alpha}(\tau_0,\Sigma) = \frac{1}{N}\text{tr}\left(\Sigma P_C(\tau_0)\overline{\overline{\mathcal{Q}}}'(\tau_0)\Sigma\overline{\overline{T}}(\tau_0)P_C(\tau_0)\right),$$
$$\text{Bias}^{\ddagger}_{\lambda,\tau\tau}(\tau_0,\Sigma) = \frac{1}{N}\text{tr}\left(\Sigma P_C(\tau_0)\overline{\overline{\mathcal{Q}}}'(\tau_0)\Sigma\overline{\overline{\mathcal{Q}}}(\tau_0)P_C(\tau_0)\right).$$

Proof. See Section C.6 in the web appendix. □

Using Lemma A.7 in the web appendix, we can show that $\text{Bias}^{\ddagger}_{\lambda}(\tau_0,\Sigma) = O(1/\min\{n_1,\ldots,n_R\})$. As in the case of homoskedastic setting, the bias becomes negligible if all group sizes are large. However, in settings with some fixed group sizes, the bias correction is necessary for valid inference. It is clear that the plug-in estimator of $\Omega^{\ddagger}(\omega_0,\lambda_0,\Sigma)$ also requires consistent estimators for the terms involving Σ. Since $\tilde{V} = \tilde{V}(\beta_0,\zeta_0) = Q_C(\tau_0)V$, we have $\text{E}(\tilde{V}\odot\tilde{V}) = (Q_C(\tau_0)\odot Q_C(\tau_0))\,(\sigma_1^2,\ldots,\sigma_n^2)'$, where $\odot$ denotes the Hadamard product. This implies an estimator for $(\sigma_1^2,\ldots,\sigma_n^2)'$ as

$$(\widehat{\sigma}_1^2,\ldots,\widehat{\sigma}_n^2)' = \left(Q_C(\widehat{\tau}^{\ddagger})\odot Q_C(\widehat{\tau}^{\ddagger})\right)^{-}(\widehat{V}\odot\widehat{V}), \tag{4.35}$$

where A^{-} denotes the generalized inverse of A and $\widehat{V} = \widehat{V}(\widehat{\beta}^{\ddagger},\widehat{\zeta}^{\ddagger}) = Q_C(\widehat{\tau}^{\ddagger})e^{\widehat{\tau}^{\ddagger}M}(e^{\widehat{\alpha}^{\ddagger}W}Y - Z\widehat{\beta}^{\ddagger})$. Note that elements of $\Omega^{\ddagger}(\omega_0)$ takes forms of either $\text{tr}(\Sigma C)$ or $\text{tr}(\Sigma A\Sigma B)$. In the next theorem, we show the effect of replacing the unknown Σ by $\widehat{\Sigma} = \text{Diag}(\widehat{\sigma}_1^2,\ldots,\widehat{\sigma}_n^2)$ in $\text{tr}(\Sigma C)$ and $\text{tr}(\Sigma A\Sigma B)$.

Theorem 4.7. *Assume* $\Pi(\tau) = (Q_C(\tau)\odot Q_C(\tau))^{-1}$ *exists for* τ *in a neighborhood of* τ_0 *and is bounded in row and column sum norms. Let* A *and* B *be two* $n\times n$ *matrices that are bounded in row and column sum norms with zero diagonal elements. Let* C *be an* $n\times n$ *matrix that has uniformly bounded diagonal elements. Then,*

(i) $\frac{1}{N}\text{tr}\left(\widehat{\Sigma}C\right) - \frac{1}{N}\text{tr}\left(\Sigma C\right) = o_p(1)$,

(ii) $\frac{1}{N}\text{tr}\left(\widehat{\Sigma}A\widehat{\Sigma}B\right) - \frac{1}{N}\text{tr}\left(\Sigma A\Sigma B\right) - \frac{2}{N}\text{tr}\left((A\odot B)\Pi(\tau_0)\Lambda(\Sigma)\Pi(\tau_0)\right) = o_p(1)$,

where $\Lambda(\Sigma) = (Q_C(\tau_0)\Sigma Q_C(\tau_0)) \odot (Q_C(\tau_0)\Sigma Q_C(\tau_0))$.

Proof. See Section C.7 in the web appendix. □

Our results in Theorems 4.6 and 4.7 suggest that the bias corrected estimator of $\Omega^\ddagger(\omega_0)$ is given by

$$\widehat{\Omega}^\ddagger = \Omega^\ddagger(\widehat{\omega}^\ddagger, \widehat{\lambda}^\ddagger, \widehat{\Sigma}) - \text{Bias}^\ddagger_\lambda(\widehat{\tau}^\ddagger, \widehat{\Sigma}) - \text{Bias}^\ddagger_\Sigma(\widehat{\tau}^\ddagger, \widehat{\Sigma}). \tag{4.36}$$

The first term $\text{Bias}^\ddagger_\lambda(\widehat{\tau}^\ddagger, \widehat{\Sigma})$ is the plug-in estimator of $\text{Bias}^\ddagger_\lambda(\tau_0, \Sigma)$ given in Theorem 4.6. The second term $\text{Bias}^\ddagger_\Sigma(\widehat{\tau}^\ddagger, \widehat{\Sigma})$ arises because of replacing Σ with $\widehat{\Sigma}$ and is an $(k+2)\times(k+2)$ matrix given as

$$\text{Bias}^\ddagger_\Sigma(\widehat{\tau}^\ddagger, \widehat{\Sigma}) = \begin{pmatrix} 0_{k\times k} & 0_{k\times 1} & 0_{k\times 1} \\ 0_{1\times k} & \text{Bias}^\ddagger_{\Sigma,\alpha\alpha}(\widehat{\tau}^\ddagger, \widehat{\Sigma}) & \text{Bias}^\ddagger_{\Sigma,\alpha\tau}(\widehat{\tau}^\ddagger, \widehat{\Sigma}) \\ 0_{1\times k} & \text{Bias}^\ddagger_{\Sigma,\tau\alpha}(\widehat{\tau}^\ddagger, \widehat{\Sigma}) & \text{Bias}^\ddagger_{\Sigma,\tau\tau}(\widehat{\tau}^\ddagger, \widehat{\Sigma}) \end{pmatrix},$$

where

$$\text{Bias}^\ddagger_{\Sigma,\alpha\alpha}(\widehat{\tau}^\ddagger, \widehat{\Sigma}) = \frac{2}{N}\text{tr}\left(\left(\overline{\overline{T}}(\widehat{\tau}^\ddagger) \odot \overline{\overline{T}}^s(\widehat{\tau}^\ddagger) - P_C(\widehat{\tau}^\ddagger)\overline{\overline{T}}'(\widehat{\tau}^\ddagger) \odot \overline{\overline{T}}(\widehat{\tau}^\ddagger)P_C(\widehat{\tau}^\ddagger)\right)\Pi(\widehat{\tau}^\ddagger)\Lambda(\widehat{\Sigma})\Pi(\widehat{\tau}^\ddagger)\right),$$
$$\text{Bias}^\ddagger_{\Sigma,\alpha\tau}(\widehat{\tau}^\ddagger, \widehat{\Sigma}) = \frac{2}{N}\text{tr}\left(\left(\overline{\overline{T}}(\widehat{\tau}^\ddagger) \odot \overline{\overline{Q}}^s(\widehat{\tau}^\ddagger) - P_C(\widehat{\tau}^\ddagger)\overline{\overline{Q}}'(\widehat{\tau}^\ddagger) \odot \overline{\overline{T}}(\tau_0)P_C(\widehat{\tau}^\ddagger)\right)\Pi(\widehat{\tau}^\ddagger)\Lambda(\widehat{\Sigma})\Pi(\widehat{\tau}^\ddagger)\right),$$
$$\text{Bias}^\ddagger_{\Sigma,\tau\tau}(\widehat{\tau}^\ddagger, \widehat{\Sigma}) = \frac{2}{N}\text{tr}\left(\left(\overline{\overline{Q}}(\widehat{\tau}^\ddagger) \odot \overline{\overline{Q}}^s(\widehat{\tau}^\ddagger) - P_C(\widehat{\tau}^\ddagger)\overline{\overline{Q}}'(\widehat{\tau}^\ddagger) \odot \overline{\overline{Q}}(\widehat{\tau}^\ddagger)P_C(\widehat{\tau}^\ddagger)\right)\Pi(\widehat{\tau}^\ddagger)\Lambda(\widehat{\Sigma})\Pi(\widehat{\tau}^\ddagger)\right),$$

and $\Lambda(\widehat{\Sigma}) = \left(Q_C(\widehat{\tau}^\ddagger)\widehat{\Sigma}Q_C(\widehat{\tau}^\ddagger)\right) \odot \left(Q_C(\widehat{\tau}^\ddagger)\widehat{\Sigma}Q_C(\widehat{\tau}^\ddagger)\right)$.

5 Simulations

In this section, we investigate the finite sample properties of our proposed estimation and inference methodologies through an extensive simulation study. The data generating process takes the following form:

$$Y_r = e^{-\alpha_0 W_r}\left(X_{1r}\beta_{10} + X_{2r}\beta_{20} + W_r X_{1r}\beta_{30} + W_r X_{2r}\beta_{40} + \lambda_{r0}\mathbf{1}_{n_r} + e^{-\tau_0 M_r}V_r\right), \tag{5.1}$$

for $r = 1, 2, \ldots, R$, where X_{1r} and X_{2r} are the $n_r \times 1$ vectors of observed characteristics whose elements are independently generated from the uniform distribution $U(1, 5)$. The associated parameters are set to $(\beta_{10}, \beta_{20}, \beta_{30}, \beta_{40})' = (1.2, 0.6, -0.4, 0.1)'$. For the endogenous and correlated effects, we consider $\alpha_0 \in \{-2, 0, 2\}$ and $\tau_0 \in \{-1, 0, 1\}$. We generate the group fixed effects, λ_{r0}'s, independently from the standard normal distribution.

We set $R = 30$ and allow n_r to vary across these 30 groups by randomly assigning a value from the set of integers $15, 16, \ldots, 20$ to each group size. Therefore, the total number of observations can vary between 450 and 600. Following Liu and Lee (2010), we generate W_r in two steps. First, we draw an integer value ϑ_{ir} uniformly from the set $1, 2, 3, 4, 5$. Then, if $\vartheta_{ir} + i \leq n_r$, the $(i+1)$th, $\ldots$, $(i + \vartheta_{ir})$th elements of the ith row of W_r are set to one, and the rest of the elements in the ith row are set to zero. On the other hand, if $\vartheta_{ir} + i > n_r$, the first $(\vartheta_{ir} + i - n_r)$ entries of the ith row are set to one, and the others are set to zero. We row-normalize W_r generated in this way and set $W = \text{Blkdiag}\big(W_1, \ldots, W_R\big)$. We generate $M = \text{Blkdiag}\big(M_1, \ldots, M_R\big)$ using the same method but a different random number generator, ensuring that $M \neq W$.

We specify the disturbance terms as $v_{ir} = \omega_{ir}\varepsilon_{ir}$, where ω_{ir}'s are the variance terms and ε_{ir}'s are independent and identically distributed random variables with a mean of 0 and a variance of 1. We consider three cases for the distribution of ε_{ir}: (i) $\varepsilon_{ir} \sim N(0,1)$, (ii) $\varepsilon_{ir} \sim \text{Gamma}(2,1)$, where $\text{Gamma}(a,b)$ is the standardized gamma distribution with shape and scale parameters a and b, and (iii) $\varepsilon_{ir} \sim \chi^2_3$, where χ^2_ν is the standardized chi-squared distribution with ν degrees of freedom.

In the homoskedastic case, we set $\omega_{ir} = 1$ in all scenarios. In the heteroskedastic case, we consider three scenarios for ω_{ir}. In the first scenario (Case 1), if the number of peers for the ith entity in the rth group is smaller than or equal to the average number of peers in that group, ω_{ir} is set to the number of peers. Otherwise, it is set to the square of the inverse of the number of peers. Thus, this scenario allows heteroskedasticity to vary across i and r. In the second scenario (Case 2), if the group size for the rth group is greater than the average group size in the sample, ω_{ir} is set to the group size. Otherwise, it is set to the square of the inverse of the group size. Therefore, heteroskedasticity varies across groups but does not vary within a group. In the last scenario (Case 3), we set ω_{ir} to $n_r(|X_{1,ir}|+|X_{2,ir}|)/\sum_{j=1}^{n_r}(|X_{1,jr}|+|X_{2,jr}|)$, allowing heteroskedasticity to vary across i and r. In all heteroskedastic scenarios, we only consider $\varepsilon_{ir} \sim N(0,1)$.

We set the number of repetitions to 1000 in all cases. We report bias, empirical standard deviation, average estimated asymptotic standard error, and empirical coverage ratio (at the 5% significance level). We present all simulation results in 16 tables. We consider the simulation results in Tables 1–4 and leave the rest to Section D in the web appendix. We summarize the salient features of the simulation results below.

1. Table 1 presents the simulation results for the QMLE of α_0, τ_0, and β_{10} under homoskedasticity. The first, second, and third panels illustrate the performance of $\widehat{\alpha}$, $\widehat{\tau}$, and $\widehat{\beta}_1$, respectively. The QMLE reports negligible bias for all parameters, and its performance does not depend on the type of error distributions. The average estimated

asymptotic standard errors closely match the corresponding empirical standard deviations in all cases. Consequently, the empirical coverage ratios are close to the nominal value of 95% in all cases. Our results also show that the QMLE performs satisfactorily for β_2, β_3 and β_4 (see Table D.1 in the web appendix). The simulation results for the heteroskedastic cases are presented in Tables D.2 and D.3 in the web appendix. These results reveal that although the QMLE performs satisfactorily in terms of bias, it reports empirical coverage rates that are significantly below 95%, especially in the case of α_0 and τ_0. Overall, consistent with our theoretical results, the QMLE performs satisfactorily under homoskedasticity. However, under heteroskedasticity, it may report smaller estimated asymptotic standard errors than the corresponding empirical standard deviations.

2. Table 2 presents the simulation results for the RME based on the transformation approach under heteroskedasticity. Overall, these results show that the RME performs satisfactorily for α_0, τ_0, and β_{10} in all cases, except for some instances in the first two panels, where the empirical coverage rates fall slightly below 95% level. Its performance for β_{20}, β_{30}, and β_{40} is excellent (see Table D.6 in the web appendix). We also investigate the performance of the RME under the homoskedastic cases, and these results are presented in Tables D.4 and D.5 in the web appendix. These results also indicate that the RME performs satisfactorily.

3. The simulation results for the ME based on the direct approach under homoskedasticity are presented in Table 3. In terms of bias, the ME performs satisfactorily in all cases. It reports empirical coverage rates that are close to the nominal value of 95% in all cases. Its performance for β_{20}, β_{30} and β_{40} is excellent as can be seen from the results in Table D.7 in the web appendix. The simulation results for the heteroskedastic cases are presented in Tables D.8 and D.9 in the web appendix. These results indicate that its performance in terms of bias is not affected by heteroskedasticity. However, it reports empirical coverage rates significantly below the 95% nominal level in the cases of α_0 and τ_0 (see Table D.8).

4. Finally, we consider the simulation results for the RME based on the direct approach. The results in Table 4 demonstrate that this estimator performs satisfactorily under all heteroskedastic cases. Its performance for β_{20}, β_{30}, and β_{40} under heteroskedasticity is evident from the simulation results presented in Table D.12 in the web appendix, which also show satisfactory performance. The simulation results for the homoskedastic cases are displayed in Tables D.10 and D.11 in the web appendix. Overall, these results indicate that the RME performs well in all homoskedastic cases as well.

Table 1: Transformation approach: The QMLE results for $\widehat{\alpha}$, $\widehat{\tau}$ and $\widehat{\beta}_1$ under homoskedasticity

	Normal			Gamma			Chi-squared		
	$\tau_0 = -1$	$\tau_0 = 0$	$\tau_0 = 1$	$\tau_0 = -1$	$\tau_0 = 0$	$\tau_0 = 1$	$\tau_0 = -1$	$\tau_0 = 0$	$\tau_0 = 1$
Results for $\widehat{\alpha}$									
$\alpha_0 = -2$	0.0070	0.0013	0.0015	0.0019	0.0032	−0.0003	0.0054	0.0030	0.0025
	0.0585	*0.0651*	*0.0393*	*0.0602*	*0.0688*	*0.0402*	*0.0577*	*0.0675*	*0.0377*
	0.0576	*0.0653*	*0.0377*	*0.0575*	*0.0655*	*0.0386*	*0.0575*	*0.0654*	*0.0385*
	0.939	**0.947**	**0.938**	**0.934**	**0.938**	**0.949**	**0.945**	**0.943**	**0.958**
$\alpha_0 = 0$	0.0041	0.0056	−0.0005	0.0033	0.0015	0.0011	0.0057	0.0060	0.0040
	0.0594	*0.0678*	*0.0378*	*0.0566*	*0.0681*	*0.0395*	*0.0584*	*0.0703*	*0.0412*
	0.0575	*0.0657*	*0.0376*	*0.0574*	*0.0653*	*0.0385*	*0.0576*	*0.0656*	*0.0385*
	0.942	**0.937**	**0.951**	**0.953**	**0.937**	**0.943**	**0.934**	**0.932**	**0.934**
$\alpha_0 = 2$	0.0043	0.0066	0.0000	0.0046	0.0004	0.0028	0.0032	0.0058	0.0013
	0.0566	*0.0694*	*0.0386*	*0.0589*	*0.0649*	*0.0386*	*0.0610*	*0.0695*	*0.0405*
	0.0575	*0.0655*	*0.0377*	*0.0576*	*0.0654*	*0.0386*	*0.0575*	*0.0656*	*0.0385*
	0.955	**0.928**	**0.941**	**0.943**	**0.948**	**0.956**	**0.932**	**0.938**	**0.936**
Results for $\widehat{\tau}$									
$\alpha_0 = -2$	−0.0120	0.0034	0.0097	−0.0060	−0.0011	0.0098	−0.0120	0.0012	0.0063
	0.0758	*0.0753*	*0.0644*	*0.0728*	*0.0756*	*0.0662*	*0.0780*	*0.0756*	*0.0666*
	0.0731	*0.0729*	*0.0638*	*0.0730*	*0.0731*	*0.0643*	*0.0730*	*0.0731*	*0.0642*
	0.935	**0.939**	**0.938**	**0.949**	**0.947**	**0.933**	**0.924**	**0.943**	**0.938**
$\alpha_0 = 0$	−0.0081	0.0007	0.0130	−0.0090	0.0004	0.0095	−0.0085	−0.0028	0.0059
	0.0776	*0.0746*	*0.0656*	*0.0758*	*0.0739*	*0.0684*	*0.0759*	*0.0782*	*0.0674*
	0.0730	*0.0731*	*0.0637*	*0.0730*	*0.0731*	*0.0643*	*0.0730*	*0.0732*	*0.0642*
	0.933	**0.949**	**0.945**	**0.934**	**0.948**	**0.936**	**0.941**	**0.947**	**0.940**
$\alpha_0 = 2$	−0.0101	0.0027	0.0081	−0.0051	0.0008	0.0079	−0.0081	−0.0026	0.0054
	0.0737	*0.0750*	*0.0666*	*0.0756*	*0.0735*	*0.0654*	*0.0781*	*0.0762*	*0.0656*
	0.0730	*0.0730*	*0.0638*	*0.0730*	*0.0731*	*0.0643*	*0.0730*	*0.0732*	*0.0642*
	0.952	**0.944**	**0.942**	**0.940**	**0.946**	**0.952**	**0.919**	**0.937**	**0.942**
Results for $\widehat{\beta}_1$									
$\alpha_0 = -2$	0.0000	−0.0005	0.0005	0.0004	0.0002	−0.0017	0.0000	−0.0006	0.0001
	0.0283	*0.0320*	*0.0276*	*0.0269*	*0.0311*	*0.0274*	*0.0277*	*0.0318*	*0.0277*
	0.0274	*0.0316*	*0.0273*	*0.0273*	*0.0315*	*0.0272*	*0.0274*	*0.0315*	*0.0274*
	0.942	**0.942**	**0.949**	**0.958**	**0.950**	**0.946**	**0.943**	**0.947**	**0.948**
$\alpha_0 = 0$	0.0012	0.0003	−0.0006	−0.0011	−0.0002	0.0005	−0.0001	0.0000	0.0009
	0.0273	*0.0324*	*0.0277*	*0.0284*	*0.0310*	*0.0274*	*0.0277*	*0.0334*	*0.0281*
	0.0274	*0.0316*	*0.0273*	*0.0274*	*0.0315*	*0.0272*	*0.0274*	*0.0315*	*0.0273*
	0.949	**0.939**	**0.952**	**0.939**	**0.953**	**0.940**	**0.949**	**0.934**	**0.937**
$\alpha_0 = 2$	−0.0001	0.0008	0.0007	−0.0004	−0.0001	−0.0011	0.0008	0.0008	−0.0018
	0.0266	*0.0322*	*0.0268*	*0.0283*	*0.0321*	*0.0277*	*0.0275*	*0.0318*	*0.0278*
	0.0274	*0.0315*	*0.0273*	*0.0275*	*0.0316*	*0.0273*	*0.0273*	*0.0315*	*0.0274*
	0.957	**0.946**	**0.964**	**0.943**	**0.949**	**0.944**	**0.939**	**0.938**	**0.942**

Notes: In each (α_0, τ_0) combination, the first row gives the bias, the second row the *empirical standard deviation*, the third row the *average asymptotic standard error*, and the last row the **95% empirical coverage rate**.

Table 2: Transformation approach: The RME results for $\widehat{\alpha}$, $\widehat{\tau}$ and $\widehat{\beta}_1$ under heteroskedasticity

	Case 1			Case 2			Case 3		
	$\tau_0 = -1$	$\tau_0 = 0$	$\tau_0 = 1$	$\tau_0 = -1$	$\tau_0 = 0$	$\tau_0 = 1$	$\tau_0 = -1$	$\tau_0 = 0$	$\tau_0 = 1$
Results for $\widehat{\alpha}$									
$\alpha_0 = -2$	−0.0013	−0.0012	0.0017	0.0026	0.0032	0.0001	0.0003	0.0036	0.0016
	0.0692	*0.0926*	*0.0627*	*0.0953*	*0.1161*	*0.0798*	*0.0652*	*0.0837*	*0.0564*
	0.0681	*0.0870*	*0.0551*	*0.0873*	*0.1108*	*0.0728*	*0.0619*	*0.0772*	*0.0522*
	0.949	**0.933**	**0.923**	**0.929**	**0.944**	**0.929**	**0.935**	**0.924**	**0.935**
$\alpha_0 = 0$	−0.0032	0.0047	−0.0023	0.0033	−0.0049	0.0022	0.0018	0.0000	−0.0017
	0.0681	*0.0894*	*0.0574*	*0.0911*	*0.1145*	*0.0829*	*0.0646*	*0.0813*	*0.0565*
	0.0684	*0.0869*	*0.0543*	*0.0880*	*0.1104*	*0.0734*	*0.0617*	*0.0769*	*0.0523*
	0.953	**0.944**	**0.936**	**0.947**	**0.934**	**0.921**	**0.937**	**0.936**	**0.929**
$\alpha_0 = 2$	−0.0034	−0.0006	0.0017	0.0032	−0.0015	0.0042	0.0016	−0.0011	0.0040
	0.0683	*0.0888*	*0.0588*	*0.0918*	*0.1120*	*0.0755*	*0.0634*	*0.0808*	*0.0570*
	0.0689	*0.0869*	*0.0559*	*0.0878*	*0.1099*	*0.0738*	*0.0617*	*0.0777*	*0.0522*
	0.956	**0.948**	**0.933**	**0.930**	**0.929**	**0.940**	**0.941**	**0.937**	**0.922**
Results for $\widehat{\tau}$									
$\alpha_0 = -2$	−0.0050	0.0038	0.0126	−0.0050	0.0003	0.0140	−0.0058	0.0015	0.0101
	0.0913	*0.0955*	*0.0913*	*0.1077*	*0.1133*	*0.1039*	*0.0794*	*0.0838*	*0.0768*
	0.0875	*0.0921*	*0.0855*	*0.1026*	*0.1074*	*0.0961*	*0.0767*	*0.0793*	*0.0742*
	0.949	**0.934**	**0.929**	**0.943**	**0.933**	**0.910**	**0.944**	**0.932**	**0.943**
$\alpha_0 = 0$	−0.0020	0.0023	0.0112	−0.0070	0.0042	0.0141	−0.0059	0.0013	0.0123
	0.0919	*0.0980*	*0.0923*	*0.1066*	*0.1122*	*0.1065*	*0.0773*	*0.0890*	*0.0776*
	0.0876	*0.0918*	*0.0857*	*0.1026*	*0.1069*	*0.0970*	*0.0768*	*0.0794*	*0.0738*
	0.936	**0.931**	**0.933**	**0.941**	**0.937**	**0.923**	**0.950**	**0.908**	**0.935**
$\alpha_0 = 2$	−0.0040	0.0043	0.0140	−0.0059	0.0043	0.0094	−0.0022	0.0060	0.0081
	0.0901	*0.0970*	*0.0907*	*0.1104*	*0.1132*	*0.1047*	*0.0802*	*0.0832*	*0.0780*
	0.0881	*0.0918*	*0.0865*	*0.1027*	*0.1069*	*0.0966*	*0.0770*	*0.0790*	*0.0742*
	0.950	**0.918**	**0.931**	**0.922**	**0.929**	**0.921**	**0.950**	**0.927**	**0.925**
Results for $\widehat{\beta}_1$									
$\alpha_0 = -2$	−0.0004	−0.0007	0.0008	−0.0003	0.0010	0.0000	−0.0003	0.0015	−0.0005
	0.0295	*0.0337*	*0.0279*	*0.0277*	*0.0340*	*0.0297*	*0.0314*	*0.0398*	*0.0323*
	0.0279	*0.0320*	*0.0271*	*0.0275*	*0.0320*	*0.0286*	*0.0306*	*0.0377*	*0.0314*
	0.940	**0.933**	**0.940**	**0.944**	**0.943**	**0.939**	**0.941**	**0.930**	**0.939**
$\alpha_0 = 0$	0.0006	−0.0007	0.0010	−0.0018	−0.0009	−0.0006	0.0003	−0.0004	−0.0012
	0.0288	*0.0330*	*0.0277*	*0.0283*	*0.0327*	*0.0301*	*0.0331*	*0.0413*	*0.0329*
	0.0278	*0.0319*	*0.0272*	*0.0275*	*0.0319*	*0.0285*	*0.0306*	*0.0374*	*0.0312*
	0.939	**0.944**	**0.955**	**0.942**	**0.940**	**0.940**	**0.936**	**0.931**	**0.936**
$\alpha_0 = 2$	−0.0018	−0.0001	−0.0007	0.0007	0.0002	−0.0002	0.0001	0.0005	0.0009
	0.0291	*0.0322*	*0.0287*	*0.0288*	*0.0348*	*0.0293*	*0.0322*	*0.0386*	*0.0329*
	0.0279	*0.0320*	*0.0274*	*0.0277*	*0.0321*	*0.0287*	*0.0306*	*0.0373*	*0.0313*
	0.938	**0.950**	**0.939**	**0.942**	**0.928**	**0.952**	**0.935**	**0.944**	**0.940**

Notes: In each (α_0, τ_0) combination, the first row gives the bias, the second row the *empirical standard deviation*, the third row the *average asymptotic standard error*, and the last row the **95% empirical coverage rate**.

Table 3: Direct approach: The ME results for $\widehat{\alpha}$, $\widehat{\tau}$ and $\widehat{\beta}_1$ under homoskedasticity

	Normal			Gamma			Chi-squared		
	$\tau_0 = -1$	$\tau_0 = 0$	$\tau_0 = 1$	$\tau_0 = -1$	$\tau_0 = 0$	$\tau_0 = 1$	$\tau_0 = -1$	$\tau_0 = 0$	$\tau_0 = 1$
Results for $\widehat{\alpha}$									
$\alpha_0 = -2$	0.0069	0.0018	0.0012	0.0017	0.0032	−0.0020	0.0054	0.0034	0.0017
	0.0586	*0.0654*	*0.0516*	*0.0603*	*0.0696*	*0.0538*	*0.0576*	*0.0685*	*0.0523*
	0.0595	*0.0674*	*0.0521*	*0.0588*	*0.0671*	*0.0528*	*0.0595*	*0.0676*	*0.0529*
	0.947	**0.950**	**0.949**	**0.943**	**0.934**	**0.942**	**0.945**	**0.943**	**0.946**
$\alpha_0 = 0$	0.0041	0.0056	0.0001	0.0034	0.0015	0.0007	0.0054	0.0057	0.0041
	0.0595	*0.0680*	*0.0529*	*0.0569*	*0.0684*	*0.0520*	*0.0588*	*0.0703*	*0.0546*
	0.0597	*0.0674*	*0.0528*	*0.0589*	*0.0673*	*0.0526*	*0.0601*	*0.0677*	*0.0532*
	0.949	**0.940**	**0.947**	**0.953**	**0.936**	**0.953**	**0.945**	**0.940**	**0.943**
$\alpha_0 = 2$	0.0042	0.0069	0.0011	0.0047	0.0005	0.0035	0.0031	0.0062	−0.0005
	0.0570	*0.0702*	*0.0538*	*0.0590*	*0.0650*	*0.0512*	*0.0610*	*0.0698*	*0.0527*
	0.0596	*0.0683*	*0.0524*	*0.0600*	*0.0672*	*0.0529*	*0.0595*	*0.0676*	*0.0526*
	0.951	**0.932**	**0.943**	**0.940**	**0.959**	**0.965**	**0.941**	**0.931**	**0.950**
Results for $\widehat{\tau}$									
$\alpha_0 = -2$	−0.0118	0.0033	0.0085	−0.0053	−0.0009	0.0101	−0.0118	0.0014	0.0054
	0.0759	*0.0758*	*0.0673*	*0.0730*	*0.0750*	*0.0701*	*0.0782*	*0.0762*	*0.0698*
	0.0756	*0.0752*	*0.0696*	*0.0748*	*0.0753*	*0.0704*	*0.0754*	*0.0756*	*0.0703*
	0.944	**0.943**	**0.958**	**0.951**	**0.943**	**0.942**	**0.926**	**0.948**	**0.950**
$\alpha_0 = 0$	−0.0079	0.0010	0.0109	−0.0086	0.0010	0.0080	−0.0081	−0.0020	0.0042
	0.0776	*0.0747*	*0.0713*	*0.0759*	*0.0742*	*0.0718*	*0.0762*	*0.0785*	*0.0707*
	0.0756	*0.0760*	*0.0701*	*0.0748*	*0.0755*	*0.0700*	*0.0756*	*0.0759*	*0.0706*
	0.939	**0.958**	**0.949**	**0.943**	**0.955**	**0.941**	**0.947**	**0.934**	**0.947**
$\alpha_0 = 2$	−0.0097	0.0027	0.0056	−0.0048	0.0019	0.0056	−0.0077	−0.0025	0.0047
	0.0740	*0.0756*	*0.0730*	*0.0755*	*0.0735*	*0.0686*	*0.0782*	*0.0760*	*0.0685*
	0.0754	*0.0761*	*0.0698*	*0.0754*	*0.0756*	*0.0698*	*0.0755*	*0.0753*	*0.0701*
	0.949	**0.948**	**0.929**	**0.949**	**0.949**	**0.949**	**0.932**	**0.946**	**0.953**
Results for $\widehat{\beta}_1$									
$\alpha_0 = -2$	0.0000	−0.0005	0.0005	0.0004	0.0002	−0.0017	0.0000	−0.0006	0.0002
	0.0283	*0.0320*	*0.0278*	*0.0269*	*0.0311*	*0.0276*	*0.0277*	*0.0318*	*0.0276*
	0.0275	*0.0317*	*0.0276*	*0.0273*	*0.0316*	*0.0275*	*0.0275*	*0.0316*	*0.0275*
	0.943	**0.939**	**0.950**	**0.957**	**0.955**	**0.940**	**0.952**	**0.946**	**0.946**
$\alpha_0 = 0$	0.0012	0.0003	−0.0006	−0.0011	−0.0002	0.0006	−0.0001	0.0000	0.0009
	0.0273	*0.0324*	*0.0279*	*0.0284*	*0.0311*	*0.0276*	*0.0277*	*0.0334*	*0.0284*
	0.0276	*0.0317*	*0.0276*	*0.0274*	*0.0316*	*0.0275*	*0.0275*	*0.0316*	*0.0276*
	0.943	**0.941**	**0.951**	**0.942**	**0.955**	**0.942**	**0.949**	**0.937**	**0.940**
$\alpha_0 = 2$	−0.0001	0.0008	0.0009	−0.0004	−0.0001	−0.0009	0.0008	0.0008	−0.0018
	0.0266	*0.0322*	*0.0269*	*0.0283*	*0.0321*	*0.0278*	*0.0275*	*0.0318*	*0.0279*
	0.0276	*0.0317*	*0.0276*	*0.0277*	*0.0318*	*0.0275*	*0.0276*	*0.0316*	*0.0277*
	0.959	**0.944**	**0.967**	**0.939**	**0.946**	**0.942**	**0.946**	**0.942**	**0.933**

Notes: In each (α_0, τ_0) combination, the first row gives the bias, the second row the *empirical standard deviation*, the third row the *average asymptotic standard error*, and the last row the **95% empirical coverage rate**.

Table 4: Direct approach: The RME results for $\widehat{\alpha}$, $\widehat{\tau}$ and $\widehat{\beta}_1$ under heteroskedasticity

	Case 1			Case 2			Case 3		
	$\tau_0 = -1$	$\tau_0 = 0$	$\tau_0 = 1$	$\tau_0 = -1$	$\tau_0 = 0$	$\tau_0 = 1$	$\tau_0 = -1$	$\tau_0 = 0$	$\tau_0 = 1$
Results for $\widehat{\alpha}$									
$\alpha_0 = -2$	0.0033	0.0046	0.0021	0.0077	0.0082	0.0005	0.0023	0.0013	0.0026
	0.0717	*0.0828*	*0.0546*	*0.0782*	*0.0860*	*0.0602*	*0.0588*	*0.0717*	*0.0509*
	0.0713	*0.0829*	*0.0544*	*0.0762*	*0.0845*	*0.0571*	*0.0588*	*0.0677*	*0.0530*
	0.941	**0.956**	**0.947**	**0.924**	**0.942**	**0.929**	**0.951**	**0.929**	**0.955**
$\alpha_0 = 0$	0.0026	0.0026	0.0052	0.0056	0.0053	0.0039	0.0037	0.0005	−0.0016
	0.0735	*0.0858*	*0.0566*	*0.0768*	*0.0832*	*0.0589*	*0.0595*	*0.0672*	*0.0533*
	0.0709	*0.0824*	*0.0549*	*0.0760*	*0.0824*	*0.0572*	*0.0584*	*0.0674*	*0.0525*
	0.933	**0.939**	**0.951**	**0.935**	**0.940**	**0.948**	**0.942**	**0.932**	**0.943**
$\alpha_0 = 2$	−0.0032	0.0056	0.0009	0.0079	0.0087	0.0056	0.0032	0.0055	0.0023
	0.0699	*0.0810*	*0.0546*	*0.0779*	*0.0829*	*0.0602*	*0.0584*	*0.0686*	*0.0525*
	0.0718	*0.0829*	*0.0549*	*0.0763*	*0.0830*	*0.0575*	*0.0586*	*0.0675*	*0.0527*
	0.946	**0.949**	**0.941**	**0.931**	**0.947**	**0.946**	**0.940**	**0.937**	**0.957**
Results for $\widehat{\tau}$									
$\alpha_0 = -2$	−0.0055	0.0007	0.0105	−0.0104	−0.0042	0.0093	−0.0085	0.0026	0.0078
	0.0968	*0.0948*	*0.0852*	*0.1027*	*0.0969*	*0.0859*	*0.0753*	*0.0738*	*0.0687*
	0.0898	*0.0912*	*0.0844*	*0.0978*	*0.0974*	*0.0868*	*0.0759*	*0.0767*	*0.0709*
	0.924	**0.927**	**0.943**	**0.936**	**0.946**	**0.945**	**0.941**	**0.953**	**0.951**
$\alpha_0 = 0$	−0.0073	0.0011	0.0033	−0.0065	−0.0012	0.0080	−0.0079	0.0036	0.0115
	0.0938	*0.0958*	*0.0866*	*0.0975*	*0.0969*	*0.0881*	*0.0727*	*0.0758*	*0.0707*
	0.0898	*0.0916*	*0.0844*	*0.0967*	*0.0963*	*0.0871*	*0.0755*	*0.0767*	*0.0707*
	0.938	**0.927**	**0.946**	**0.938**	**0.949**	**0.942**	**0.949**	**0.955**	**0.946**
$\alpha_0 = 2$	−0.0038	0.0018	0.0113	−0.0107	−0.0044	0.0053	−0.0015	0.0020	0.0074
	0.0942	*0.0946*	*0.0814*	*0.0952*	*0.0941*	*0.0901*	*0.0761*	*0.0781*	*0.0706*
	0.0900	*0.0911*	*0.0849*	*0.0969*	*0.0965*	*0.0875*	*0.0758*	*0.0765*	*0.0709*
	0.923	**0.942**	**0.946**	**0.946**	**0.956**	**0.942**	**0.943**	**0.953**	**0.961**
Results for $\widehat{\beta}_1$									
$\alpha_0 = -2$	0.0000	−0.0009	0.0007	−0.0001	0.0015	0.0001	−0.0006	0.0013	−0.0004
	0.0298	*0.0329*	*0.0286*	*0.0278*	*0.0341*	*0.0287*	*0.0314*	*0.0395*	*0.0316*
	0.0293	*0.0330*	*0.0291*	*0.0287*	*0.0326*	*0.0281*	*0.0323*	*0.0390*	*0.0326*
	0.947	**0.944**	**0.953**	**0.953**	**0.946**	**0.940**	**0.958**	**0.946**	**0.954**
$\alpha_0 = 0$	0.0001	−0.0011	0.0012	−0.0015	−0.0005	0.0001	0.0000	−0.0004	−0.0008
	0.0299	*0.0335*	*0.0293*	*0.0292*	*0.0324*	*0.0283*	*0.0331*	*0.0414*	*0.0337*
	0.0292	*0.0331*	*0.0291*	*0.0287*	*0.0325*	*0.0280*	*0.0323*	*0.0388*	*0.0324*
	0.935	**0.941**	**0.946**	**0.944**	**0.948**	**0.945**	**0.942**	**0.944**	**0.929**
$\alpha_0 = 2$	−0.0014	0.0007	−0.0006	0.0009	0.0006	−0.0003	0.0002	0.0006	0.0014
	0.0289	*0.0325*	*0.0295*	*0.0298*	*0.0344*	*0.0283*	*0.0332*	*0.0387*	*0.0330*
	0.0295	*0.0330*	*0.0289*	*0.0288*	*0.0327*	*0.0281*	*0.0323*	*0.0386*	*0.0325*
	0.956	**0.952**	**0.952**	**0.947**	**0.927**	**0.954**	**0.940**	**0.953**	**0.952**

Notes: In each (α_0, τ_0) combination, the first row gives the bias, the second row the *empirical standard deviation*, the third row the *average asymptotic standard error*, and the last row the **95% empirical coverage rate**.

6 Application to the Add Health data

In this section, we use the Add Health data to study the effects of peers on academic achievement, participation in recreational activities, and smoking. The Add Health data sets were collected in three waves from adolescents in grades 7 to 12 across 132 public and private schools and include variables on demographic background, academic performance, health-related behaviors, and friendship networks of participants. We use data from the first wave, which is an in-school survey covering 90,000 students. Our first outcome variable is the grade point average (GPA), computed from several subjects, including English or language arts, history or social science, mathematics, and science. The second variable is an index of participation in recreational activities, such as educational, artistic, and sports organizations and clubs. As in Bramoullé et al. (2009), this variable takes values from 0 to 4: if the number of recreational activities a student participates in is fewer than 4, it reflects the actual number of activities; if it is 4 or more, the value is capped at 4. The third outcome variable is smoking, measured by the number of smoking days per month.

Following Lin (2010), we formulate groups at the school-grade level, resulting in a total of 792 groups. The data include information on the five male and five female friends nominated by each respondent. We use these friendship nominations to construct the network matrix. Specifically, we set the (i, j)th element of W_r to 1 if the ith student in group r nominated the jth student as a friend; otherwise, it is set to 0. We then set $M_r = W_r$ for $r = 1, \ldots, R$. We do not row-normalize W_r because the parameter space of α and τ in our setting is not constrained to any specific interval, allowing the model to have a reduced form. This feature of our model contrasts with the spatial autoregressive network models considered in the literature (Bramoullé et al., 2009, 2020; Hsieh and Lee, 2016; Lee, 2007; Lee et al., 2010; Lin, 2010, 2015; Liu and Lee, 2010).

We consider the following explanatory variables (Boucher et al., 2024; Bramoullé et al., 2009; Hsieh and Lee, 2016; Lin, 2010, 2015): age, gender dummy variable (female is the base category), race dummy variables (white is the base category), a dummy variable indicating whether the respondent lives with both parents, dummy variables for mother's education (high-school is the base category), a dummy variable indicating whether the mother is on welfare, and the mother's occupation (staying home is the base category). We consider the same set of explanatory variables for the contextual variables. Due to missing observations and computational constraints, we focus on a sample that includes groups with more than 11 students and fewer than 80 students.

The summary statistics for the entire sample and our final sample are presented in Table 5. We observe that the first two sample moments of our final sample are close to those of the entire sample for most variables. The average GPA is around 2.91, the average participation in recreational activities is 2.23, and the average number of smoking days per month is approximately 3.34. In our sample, the mean age is 14.82; 48% are male, 24.5% are Black, 3.4% are Asian, 9.2% are Hispanic, and 6.4% are from other races. The percentage of students living with both parents is 72%. We have four categories for mother's education level, with the high-school level being the base category: 10% of respondents' mothers have less than a high-school education, 40.8% have more than a high-school education, and 9.9% of respondents' mothers' education is missing. The percentage of respondents whose mothers are on welfare is around 11%. Finally, there are four categories for the mother's occupation, with staying home being the base category: 8.7% of respondents' mothers' occupations information are missing, 25.4% of respondents' mothers have professional jobs, and 35% have other jobs.[4]

For the three outcome variables, we estimate our model in (2.9) only by the ME and RME from the direct approach. Note that since the network matrices W_r and M_r are not row-normalized, we can not use the QMLE and the RME from the transformation approach. For each outcome variable, we also report a pseudo R^2 measure computed by $R^2 = 1 - \widehat{V}'\widehat{V}/((Y-\overline{Y})'(Y-\overline{Y}))$, where $\widehat{V} = e^{\widehat{\tau}M}(e^{\widehat{\alpha}W}Y - Z\widehat{\beta} - C_{\lambda}\widehat{\lambda})$.

The estimation results for GPA are presented in Table 6. The reported pseudo R^2 measures based on the ME and RME are 0.226 and 0.225, respectively. The estimates of α from the ME and RME are -0.0110 and -0.0030, respectively, both statistically insignificant. In contrast, the estimates of τ are -0.0876 and -0.1024, respectively, and statistically significant. These results suggest that, after controlling for own effects, contextual effects, group fixed effects, correlation in unobserved factors, and heteroskedasticity in the error terms, there is no statistical evidence that a student's GPA is influenced by the sum of their peers' GPAs. However, the statistically significant estimates of τ indicate that unobserved factors remain correlated, even after accounting for these controls.

The ME and RME provide similar estimates for both own and contextual effects. For the own effects, the estimated coefficients for age and gender (male) are negative and statistically significant, indicating that older and male students tend to perform worse. The estimated coefficients for the race categories are statistically significant for Black, Asian, and Hispanic dummy variables. These estimates suggest that Black and Hispanic students score lower than White students, while Asian students score higher. Both estimators report positive and statistically significant coefficients for the dummy variable indicating whether the respondent lives with both parents, suggesting that students living with both parents

[4]There are 15 occupation categories. Following Lin (2010), we combine these categories into four categories along with the missing indicators.

perform relatively better. Regarding the mother's education dummy variables, both estimators report negative and statistically significant estimates for students whose mothers have less than a high school education, and positive, statistically significant estimates for those whose mothers have education levels above high school. This suggests that students whose mothers have less than a high school education tend to have lower GPAs, while those whose mothers have more education tend to have higher GPAs. Both estimators report positive and statistically significant estimates for the dummy variable indicating whether the respondent's mother is on welfare, suggesting that students with mothers on welfare tend to have higher GPAs. Finally, for the mother's occupation categories, both estimators report positive and statistically significant estimates for the professional job category.

Regarding the contextual effects, both estimators report statistically significant estimates for only two dummy variables: whether the respondent lives with both parents and whether the respondent's mother has less than a high school education. The estimated coefficient for living with both parents is positive, while the coefficient for the mother's education level being less than high school is negative. These results suggest that students whose peers live with both parents tend to have higher GPAs, while students whose peers' mothers have less than a high school education tend to have lower GPAs.

The estimation results on participating in recreational activities are presented in Table 7. For this outcome variable, the reported pseudo R^2 measures based on the ME and RME are 0.175 and 0.176, respectively. The estimates of the endogenous effect α are 0.0146 from the ME and 0.0228 from the RME. The first estimate is statistically insignificant, while the second is significant only at the 10% significance level. On the other hand, both estimators report negative and statistically significant estimates for τ. These results suggest that, after controlling for own effects, contextual effects, group fixed effects, correlation in unobserved factors, and heteroskedasticity in the error terms, there is no statistical evidence for endogenous peer effects on participating in recreational activities. Based on a different model specification, Bramoullé et al. (2009) also report an estimate of the endogenous effect that is only significant at the 10% significance level. On the other hand, the statistically significant estimates of τ indicate that unobserved factors affecting participation remain correlated, even after accounting for these controls.

Among the own effects, the estimated coefficients for the following variables are statistically significant: age, gender (male), mother's education less than high school, mother's education more than high school, mother's education level missing, and mother's occupation being professional. The coefficients for age and gender are both negative, indicating that older and male students participate less in recreational activities. Additionally, students whose mothers' education level is missing or less than high school also participate less, while those whose mothers have more than a high school education participate more. Furthermore, students whose mothers hold a professional occupation also participate more. Among the contextual effects, the estimated coefficients for age and Black are the only ones statistically significant. We observe that having older peers increases participation, while having Black peers reduces participation.

The estimation results for smoking are presented in Table 8. The reported pseudo R^2 measures based on the ME and RME are 0.175 and 0.175, respectively. The estimates for both α and τ reported by the ME and RME estimators are statistically significant. The respective estimates for α are 0.1393 and 0.1253, while the estimates for τ are -0.2416 and -0.2274. These results provide statistical evidence of endogenous peer effects on smoking behavior, as well as correlation of unobserved factors that influence smoking frequencies.

Among the own effects, the estimated coefficients for the age variable, Black dummy variable, Hispanic dummy variable, the dummy variable indicating whether the respondent lives with both parents, and the dummy variable for mother's education less than high school are statistically significant. Of these, the estimate for age is positive, while all others are negative. Thus, smoking frequency increases with age. Black and Hispanic students smoke relatively less than white students, and students whose mothers have less than a high school education smoke less than those whose mothers have a high school education.

Among the contextual effects, the estimated coefficients for the Black dummy variable and the dummy variable indicating whether the respondent lives with both parents are negative and statistically significant. This suggests that having Black peers and peers who live with both parents reduces smoking frequency.

Table 5: Summary statistics

	Original sample		Final sample			
Variable	Mean	SD	Mean	SD	Min	Max
GPA	2.482	1.177	2.911	0.763	0	4
Recreational activities	1.811	1.459	2.227	1.475	0	4
Smoking	3.904	9.328	3.344	8.734	0	30
Age	15.023	1.700	14.820	1.904	10	19
Male	0.494	0.500	0.480	0.500	0	1
Black	0.189	0.392	0.245	0.430	0	1
Asian	0.065	0.247	0.034	0.181	0	1
Hispanic	0.146	0.353	0.092	0.289	0	1
Other race	0.056	0.230	0.064	0.244	0	1
Live with both parents	0.725	0.447	0.720	0.449	0	1
Mom education less than HS	0.103	0.304	0.100	0.301	0	1
Mom education more than HS	0.404	0.491	0.408	0.492	0	1
Mom education missing	0.113	0.316	0.099	0.299	0	1
Mom on welfare	0.009	0.093	0.011	0.105	0	1
Mom job missing	0.095	0.293	0.087	0.282	0	1
Mom on professional job	0.257	0.437	0.254	0.435	0	1
Mom on other job	0.355	0.478	0.350	0.477	0	1

Table 6: Estimation results on GPA

	ME		RME	
	α	τ	α	τ
	−0.0110	−0.0876***	−0.0030	−0.1024***
	(.010)	(.014)	(.010)	(.014)
	Own	Contextual	Own	Contextual
Age	−0.1321***	−0.0032	−0.1309***	−0.0019
	(.015)	(.002)	(.017)	(.002)
Male	−0.1129***	−0.0022	−0.1134***	−0.0031
	(.020)	(.013)	(.020)	(.012)
Black	−0.0882**	−0.0177	−0.0898**	−0.0169
	(.036)	(.013)	(.036)	(.013)
Asian	0.1055*	0.0435	0.1027**	0.0437*
	(.056)	(.028)	(.052)	(.026)
Hispanic	−0.1407***	0.0106	−0.1405***	0.0085
	(.041)	(.022)	(.046)	(.022)
Other races	−0.0151	−0.0220	−0.0146	−0.0189
	(.040)	(.026)	(.042)	(.028)
Both parents	0.1212***	0.0391***	0.1215***	0.0409***
	(.022)	(.015)	(.023)	(.015)
Less HS	−0.0957***	−0.0671***	−0.0953***	−0.0666***
	(.033)	(.023)	(.036)	(.025)
More HS	0.1242***	0.0051	0.1231***	0.0071
	(.023)	(.014)	(.023)	(.013)
Edu miss	−0.0203	−0.0231	−0.0199	−0.0212
	(.033)	(.024)	(.035)	(.024)
Welfare	0.1978**	0.0122	0.1984**	0.0129
	(.088)	(.075)	(.079)	(.083)
Job miss	−0.0592	−0.0047	−0.0581	−0.0044
	(.036)	(.026)	(.037)	(.025)
Professional	0.0595**	0.0208	0.0604**	0.0215
	(.027)	(.018)	(.027)	(.017)
Other jobs	0.0164	0.0214	0.0167	0.0210
	(.024)	(.015)	(.024)	(.015)
Pseudo R^2	0.226		0.225	

Statistical significance at 1%, 5% and 10% levels are respectively denoted by ***, ** and *.

Table 7: Estimation results on recreational activities

	ME		RME	
	α	τ	α	τ
	0.0146	−0.0338**	0.0228*	−0.0440***
	(.014)	(.017)	(.012)	(.015)
	Own	Contexual	Own	Contexual
Age	−0.2110***	0.0066**	−0.2107***	0.0075***
	(.029)	(.003)	(.030)	(.003)
Male	−0.2010***	−0.0264	−0.2011***	−0.0271
	(.037)	(.024)	(.037)	(.023)
Black	0.0292	−0.0595**	0.0262	−0.0596**
	(.065)	(.027)	(.066)	(.028)
Asian	0.1719	−0.0253	0.1723	−0.0229
	(.107)	(.058)	(.106)	(.057)
Hispanic	0.0217	−0.0253	0.0199	−0.0252
	(.081)	(.046)	(.085)	(.047)
Other races	−0.1001	0.0890*	−0.0996	0.0867
	(.079)	(.049)	(.078)	(.052)
Both parents	0.0748*	−0.0243	0.0751	−0.0201
	(.043)	(.029)	(.045)	(.030)
Less HS	−0.2127***	−0.0045	−0.2130***	−0.0062
	(.065)	(.045)	(.068)	(.046)
More HS	0.2137***	0.0212	0.2132***	0.0256
	(.046)	(.027)	(.046)	(.026)
Edu miss	−0.2250***	−0.0324	−0.2258***	−0.0345
	(.066)	(.047)	(.068)	(.047)
Welfare	0.1535	−0.0914	0.1533	−0.0886
	(.176)	(.157)	(.193)	(.165)
Job miss	−0.0608	−0.0030	−0.0606	−0.0020
	(.071)	(.050)	(.071)	(.051)
Professional	0.1910***	0.0187	0.1921***	0.0228
	(.054)	(.035)	(.053)	(.033)
Other jobs	0.0430	0.0140	0.0434	0.0162
	(.047)	(.030)	(.048)	(.029)
Pseudo R^2	0.175		0.176	

Statistical significance at 1%, 5% and 10% levels are respectively denoted by ***, ** and *.

Table 8: Estimation results on smoking

	ME		RME	
	α	τ	α	τ
	0.1393***	−0.2416***	0.1253***	−0.2274***
	(.030)	(.029)	(.040)	(.041)
	Own	Contexual	Own	Contexual
Age	0.8507***	0.0602**	0.8562***	0.0514*
	(.167)	(.024)	(.222)	(.029)
Male	−0.2097	−0.1711	−0.2081	−0.1656
	(.231)	(.156)	(.238)	(.136)
Black	−2.6739***	−0.5662**	−2.6753***	−0.4964**
	(.421)	(.247)	(.429)	(.245)
Asian	−0.8330	−0.5100	−0.8271	−0.4860*
	(.623)	(.370)	(.593)	(.273)
Hispanic	−1.3041***	−0.1839	−1.2929**	−0.1385
	(.469)	(.288)	(.522)	(.257)
Other races	0.1079	−0.0663	0.1157	−0.0581
	(.484)	(.329)	(.549)	(.295)
Both parents	−1.2153***	−0.5759***	−1.2072***	−0.5522***
	(.263)	(.189)	(.282)	(.195)
Less HS	1.2898***	0.1930	1.2996***	0.1993
	(.391)	(.282)	(.478)	(.291)
More HS	0.0195	0.0166	0.0285	0.0346
	(.276)	(.177)	(.273)	(.168)
Edu missing	−0.0664	−0.1014	−0.0585	−0.0826
	(.406)	(.293)	(.370)	(.261)
Welfare	−0.3750	−0.1072	−0.3720	−0.1170
	(1.043)	(.838)	(1.002)	(.651)
Job missing	−0.1558	−0.0919	−0.1513	−0.0744
	(.433)	(.312)	(.418)	(.291)
Professional	−0.2074	−0.0099	−0.2029	−0.0077
	(.320)	(.212)	(.316)	(.182)
Other jobs	0.0269	−0.1198	0.0336	−0.1188
	(.283)	(.187)	(.296)	(.170)
Pseudo R^2	0.175		0.175	

Statistical significance at 1%, 5% and 10% levels are respectively denoted by ***, ** and *.

7 Conclusion

In this paper, we contributed to the peer effects literature by introducing a new model of peer effects due to socially interacting agents. The MES network model accommodates the endogenous effect, the contextual effects, heterogeneity across groups, the correlation in unobserved characteristics of members, as well as an unknown form of heteroskedasticity. We provided an underlying theoretical framework which yields the MES network model, and proposed consistent estimation and inference methods for its parameters under both homoskedastic and heteroskedastic cases. We proposed a transformation approach and a direct approach for estimation where the latter does not require row normalization of network matrices. In an extensive Monte Carlo study, we showed evidence for the satisfactory finite sample performance of the proposed estimators. Moreover, in three different empirical applications using the well-known Add Health data set, we illustrated the use of MES network model in identifying peer effects on academic success, recreational activities and smoking frequency of adolescents. Finally, we note that the MES network model can be further extended to simultaneous modeling of peer effects and network formation (Goldsmith-Pinkham and Imbens, 2013; Hsieh and Lee, 2016; Hsieh and Lin, 2021). We leave this extension to a future study.

References

Blume, Lawrence E., William A. Brock, Steven N. Durlauf, and Rajshri Jayaraman (2015). "Linear Social Interactions Models". In: *Journal of Political Economy* 123.2, pp. 444–496.

Boucher, Vincent, Michelle Rendall, Philip Ushchev, and Yves Zenou (2024). "Toward a General Theory of Peer Effects". In: *Econometrica* 92.2, pp. 543–565.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin (2009). "Identification of peer effects through social networks". In: *Journal of Econometrics* 150.1, pp. 41 –55.

— (2020). "Peer Effects in Networks: A Survey". In: *Annual Review of Economics* 12.1, pp. 603–629.

Brock, William A. and Steven N. Durlauf (2001). "Discrete Choice with Social Interactions". In: *The Review of Economic Studies* 68.2, pp. 235–260.

Calvó-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou (2009). "Peer Effects and Social Networks in Education". In: *The Review of Economic Studies* 76.4, pp. 1239–1267.

Chiu, Tom YM, Tom Leonard, and Kam-Wah Tsui (1996). "The matrix-logarithmic covariance model". In: *Journal of the American Statistical Association* 91.433, pp. 198–210.

Clark, Andrew E. and Youenn Lohéac (2007). ""It wasn't me, it was them!" Social influence in risky behavior by adolescents". In: *Journal of Health Economics* 26.4, pp. 763–784.

Debarsy, Nicolas, Fei Jin, and Lung fei Lee (2015). "Large sample properties of the matrix exponential spatial specification with an application to FDI". In: *Journal of Econometrics* 188.1.

Fruehwirth, Jane Cooley (2014). "Can Achievement Peer Effect Estimates Inform Policy? A View From Inside The Black Box". In: *The Review of Economics and Statistics* 96.3, pp. 514–523.

Goldsmith-Pinkham, Paul and Guido W. Imbens (2013). "Social Networks and the Identification of Peer Effects". In: *Journal of Business & Economic Statistics* 31.3, pp. 253–264.

Hsieh, Chih-Sheng and Lung Fei Lee (2016). "A Social Interactions Model with Endogenous Friendship Formation and Selectivity". In: *Journal of Applied Econometrics* 31.2, pp. 301–319.

Hsieh, Chih-Sheng and Xu Lin (2021). "Social interactions and social preferences in social networks". In: *Journal of Applied Econometrics* 36.2, pp. 165–189.

Jennrich, Robert I. (1969). "Asymptotic Properties of Non-Linear Least Squares Estimators". In: *The Annals of Mathematical Statistics* 40.2, pp. 633–643.

Kelejian, Harry H. and Ingmar R. Prucha (2001). "On the Asymptotic Distribution of the Moran I Test Statistic with Applications". In: *Journal of Econometrics* 104.2, pp. 219–257.

— (2010). "Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances". In: *Journal of Econometrics* 157, pp. 53–67.

Lee, Lung fei, Ji Li, and Xu Lin (2014). "Binary Choice Models Wıth Social Network Under Heterogeneous Rational Expectations". In: *The Review of Economics and Statistics* 96.3, pp. 402–417.

Lee, Lung-fei (2007). "Identification and estimation of econometric models with group interactions, contextual factors and fixed effects". In: *Journal of Econometrics* 140.2, pp. 333–374.

Lee, Lung-fei, Xiaodong Liu, and Xu Lin (2010). “Specification and estimation of social interaction models with network structures”. In: *The Econometrics Journal* 13, pp. 145–176.

Lee, Lung-Fei, Xiaodong Liu, Eleonora Patacchini, and Yves Zenou (2021). “Who is the Key Player? A Network Analysis of Juvenile Delinquency”. In: *Journal of Business & Economic Statistics* 39.3, pp. 849–857.

Lee, Lung-fei and Jihai Yu (2010). “Estimation of spatial autoregressive panel data models with fixed effects”. In: *Journal of Econometrics* 154.2, pp. 165–185.

LeSage, James and R Kelley Pace (2007). “A matrix exponential spatial specification”. In: *Journal of Econometrics* 140.1, pp. 190–214.

LeSage, James and Robert K. Pace (2009). *Introduction to Spatial Econometrics (Statistics: A Series of Textbooks and Monographs*. London: Chapman and Hall/CRC.

Lin, Xu (2010). “Identifying Peer Effects in Student Academic Achievement by Spatial Autoregressive Models with Group Unobservables”. In: *Journal of Labor Economics* 28.4, pp. 825–860.

— (2015). “Utilizing spatial autoregressive models to identify peer effects among adolescents”. In: *Empirical Economics* 49, pp. 929–960.

Liu, Xiaodong and Lung-fei Lee (2010). “GMM estimation of social interaction models with centrality”. In: *Journal of Econometrics* 159.1, pp. 99 –115.

Manski, Charles F (1993). “Identification of endogenous social effects: The reflection problem”. In: *The review of economic studies* 60.3, pp. 531–542.

Moffitt, Robert A. (2001). “Policy interventions, low-level equilibria, and social interactions”. In: *Social Dynamics*. Ed. by Steven N. Darlauf and H.Peyton Young. Cambridge, MA: MIT Press.

Patacchini, Eleonora and Yves Zenou (2009). “Juvenile Delinquency and Conformism”. In: *The Journal of Law, Economics, and Organization* 28.1, pp. 1–31.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge UK: Cambridge university press.

White, Halbert G. (1980). “A Heteroskedasticity-Consistent Covariance Matrix Estimator a Direct Test for Heteroskedasticity”. In: *Econometrica* 48.4, pp. 817–838.

Yang, Ye, Osman Doğan, and Süleyman Taşpınar (2021). “Fast estimation of matrix exponential spatial models”. In: *Journal of Spatial Econometrics* 2.9, pp. 1–50.

— (2022). “Model selection and model averaging for matrix exponential spatial models”. In: *Econometric Reviews* 41.8, pp. 827–858.

— (2024). “Estimation of Matrix Exponential Unbalanced Panel Data Models with Fixed Effects: An Application to US Outward FDI Stock”. In: *Journal of Business & Economic Statistics* 42.2, pp. 469–484.

Zenou, Yves and Eleonora Patacchini (2012). “Juvenile Delinquency and Conformism”. In: *The Journal of Law, Economics, and Organization* 28.1, pp. 1–31.