# Estimation of Matrix Exponential Unbalanced Panel Data Models with Fixed Effects: An Application to US Outward FDI Stock

Ye Yang<sup>\*</sup> Osman Doğan<sup>†</sup>

Süleyman Taşpınar<sup>‡</sup>

May 3, 2023

#### Abstract

In this paper, we consider a matrix exponential unbalanced panel data model that allows for (i) spillover effects using matrix exponential terms, (ii) unobserved heterogeneity across entities and time, and (iii) potential heteroskedasticity in the error terms across entities and time. We adopt a likelihood based direct estimation approach in which we jointly estimate the common parameters and fixed effects. To ensure that our estimator has the standard large sample properties, we show how the score functions should be suitably adjusted under both homoskedasticity and heteroskedasticity. We define our suggested estimator as the root of the adjusted score functions, and therefore our approach can be called the M-estimation approach. For inference, we suggest an analytical bias correction approach involving the sample counterpart and plug-in methods to consistently estimate the variance-covariance matrix of the suggested M-estimator. Through an extensive Monte Carlo study, we show that the suggested M-estimator has good finite sample properties. In an empirical application, we use our model to investigate the third country effects on the US outward foreign direct investment (FDI) stock at the industry level.

JEL-Classification: C10, C13, C21, C23, C15.

Keywords: MESS, Unbalanced Panel Data, M Estimation, QMLE, Spatial Panel Data Model, Heteroskedasticity, Spatial Dependence.

 $<sup>^{*}</sup>$  School of Accounting, Capital University of Economics and Business, Beijing, China, email: yeyang 557@hotmail.com

 $<sup>^\</sup>dagger Department \ of \ Economics, \ Istanbul \ Technical \ University, \ Istanbul, \ Turkey, \ email: \ osmandogan@itu.edu.tr$ 

<sup>&</sup>lt;sup>‡</sup>Department of Economics, Queens College, The City University of New York, New York, U.S.A., email: staspinar@qc.cuny.edu

## 1 Introduction

Unbalanced (or incomplete) panel data are common occurrence in empirical analysis due to early exits, late entries, dormant periods of economic activity, early recordings, and so on. This is also the case for empirical studies using spatial panel data models where missing entities at a given time period do not generate any spillover effects to their neighbors. Researchers dealing with unbalanced panels often turn them to balanced panels by sub-setting the original panels to simplify econometric analysis. This paper aims to fill the gap in available estimation and inference methods for the spatial unbalanced panel data models. We consider a general spatial specification that allows for (i) spatial dependence specified through matrix exponential terms, (ii) unobserved heterogeneity across entities and time, and (iii) potential heteroskedasticity in the error terms across entities and time. We will refer to this model as the matrix exponential unbalanced panel data model with fixed effects.

The matrix exponential spatial specification (MESS) was introduced by LeSage and Pace (2007) as an alternative to spatial autoregressive specification, and has several features that make it more convenient for estimation (Chiu et al., 1996; Leonard and Hsu, 1992). In a MESS type model, the spatial dependence in the dependent variable and/or the disturbance term is formulated through a matrix exponential term of the form  $e^{\alpha W}$ , where  $\alpha$  is a scalar spatial parameter, W is an  $n \times n$ spatial weights matrix whose elements are  $O(1/h_n)$  uniformly, and  $h_n$  is a sequence that can be bounded or divergent (Lee, 2004). Therefore, the MESS imposes an exponential rate of decay for the cross-sectional dependence. Moreover, since matrix exponential terms are always invertible, there is no need to impose any restrictions on the parameter space of the spatial parameters to ensure the existence of the reduced form of the model. Finally, the likelihood based estimation of a MESS type model has the computational advantage because the likelihood function does not involve any matrix determinant terms that need to be evaluated at each iteration during estimation. Among others, see also LeSage and Pace (2009), Debarsy et al. (2015) and Yang et al. (2021).

We consider the likelihood based estimation of the matrix exponential unbalanced panel data model with fixed effects. It is well known in the literature that the quasi maximum likelihood estimator (QMLE) of the panel data regression models with entity fixed effects suffers from the incidental parameter problem when panels are short, i.e., the time dimension is short (Neyman and Scott, 1948). With both entity and time fixed effects present, a transformation approach that can wipe out the fixed effects from the model offers a feasible estimation approach. However, this approach does not extend to the unbalanced panel data models with time varying spatial weights matrices (Lee and Yu, 2010). Also, the transformation approach requires row-normalized weights matrices, which we will not dictate in this paper. Instead of a transformation approach, we adopt a likelihood-based direct approach in which we jointly estimate the common parameters and fixed effects.

In the likelihood-based direct estimation approach, we first concentrate out the fixed effects from the quasi log-likelihood function. Then, for the remaining parameters of the model, we derive their corresponding score functions and their expectations at the true parameter vector. We show that when the number of time periods is fixed, the expectations of the score functions may not tend to zero, suggesting that the probability limit of the score functions is not zero. This result suggests that the QMLE is inconsistent unless T is large. Therefore, we adjust these scores by subtracting their respective expectations from them and then use these adjusted score functions for consistent estimation. We define our suggested estimator as the root of these adjusted score functions, and therefore our approach can be called the M-estimation as suggested by Yang (2018) and Li and Yang (2021). In the case of heteroskedasticity, we also show that the QMLE may not be consistent, and therefore we adopt a similar M-estimation strategy as in the case of the homoskedastic specification by suitably adjusting the score functions so that their probability limit becomes zero under an unknown form of heteroskedasticity.

We formally establish the consistency and the asymptotic normality of the proposed M-estimator under both homoskedastic and heteroskedastic cases. The variance-covariance matrix of the proposed M-estimator takes a sandwich form. We propose an approach based on the sample counterpart and plug-in methods for consistent estimation of the variance-covariance matrix of the proposed M-estimator. We show that the expectation of the Hessian matrix can be estimated consistently by its sample counterpart, evaluated at consistent estimates, in both homoskedastic and heteroskedastic cases. In the case of the variance of the adjusted score functions, we show that the plug-in estimator can be inconsistent, and suggest an analytical bias correction in both homoskedastic and heteroskedastic cases. In an extensive Monte Carlo study, we assess the finite sample properties of the proposed estimator and the inference method. Our results attest that the proposed estimators perform satisfactorily in terms of finite sample bias and inference.

In an empirical application, we use our model to investigate the third country effects on the US outward foreign direct investment (FDI) stock at the industry level. In the literature, there are a few noteworthy empirical studies that employed unbalanced spatial panel data models. Among others, see Antweiler (2001), Bai et al. (2015), Baltagi and Chang (1994), Baltagi et al. (2001), Davis (2002), Wansbeek and Kapteyn (1989), and Wooldridge (2019). Our unbalanced panel data cover 47 host countries across 10 industries over the period 2008–2014. Our model allows for three types of spatial interaction: (i) spatial dependence in FDI specified through a matrix exponential term (Blonigen et al., 2007; Coughlin and Segev, 2000; Debarsy et al., 2015), (ii) spatially weighted third-country determinants of FDI that are motivated by the three-factor knowledge capital model considered in Baltagi et al. (2007), and (iii) spatial dependence in the error terms specified through a matrix exponential term in order to account for the transmission of shocks across host countries. Also, our model allows for the presence of an unknown form of heteroskedasticity in the error terms of the model. Our estimation results show evidence for the presence of the third country effects on the US outward FDI stock.

Theoretically, our paper belongs to the recent expanding literature on the specification and estimation of spatial panel data models. Our paper is closely related to Lee and Yu (2010) and Meng and Yang (2021). Lee and Yu (2010) consider a direct likelihood-based estimation approach for a balanced spatial panel data model with both entity and time fixed effects, and show that the variance parameter cannot be consistently estimated when the number of time periods is fixed. Similarly, Meng and Yang (2021) consider a direct likelihood-based estimation approach for an unbalanced spatial panel data model that allows for entity and time fixed effects, as well as potential heteroskedasticity in the error terms. Following Yang (2018), they propose an M-estimator for their model and establish the large sample properties of the M-estimator. Wang and Lee (2013) consider a panel data model with random effects and spatially lagged outcome variable in which some observations on the dependent variable are missing at random. Using an imputation approach for missing observations, they propose a nonlinear least squares approach and a generalized method of moments approach for the estimation of the model. Our paper differs from these papers since we specify the spatial dependence through matrix exponential terms instead of spatial autoregressive processes. In the literature, it is well known that the quasi maximum likelihood estimator (QMLE) of spatial autoregressive models may not be consistent in the presence of an unknown form of heteroskedasticity (Kelejian and Prucha, 2010; Lin and Lee, 2010; Liu and Yang, 2015). However, in the case of MESS type models, the QMLE remains to be consistent under an unknown form of heteroskedasticity provided that the panels are balanced, the spatial weights matrices are time invariant and commute, and the model involves only entity fixed effects (Debarsy et al., 2015; Zhang et al., 2019). We show that the QMLE of the matrix exponential unbalanced panel data model with entity and time fixed effects may not be consistent. Following Yang (2018), we suggest a heteroskedasticity robust M-estimator based on suitably adjusted score functions for estimation. Our estimator has the standard large sample properties irrespective of whether the number of time periods and  $h_n$  are bounded or unbounded.

The rest of the paper is organized as follows. In Section 2, we provide the specification details of the matrix exponential unbalanced panel data model with fixed effects. In Section 3, we present the details of the M-estimation methodology for the homoskedastic specification and establish the formal results for the large sample properties of the proposed estimator. In Section 4, we extend the main ideas of the M-estimation methodology to the heteroskedastic specification and establish the formal results. In Section 5, we investigate the finite sample performance of the proposed estimators through an extensive Monte Carlo study. In Section 6, we provide our empirical application on the third country effects on the US outward FDI stock at the industry level. In Section 7, we offer concluding remarks. All technical details are left to a supplementary web appendix.

## 2 Model specification

The matrix exponential unbalanced panel data model requires stacking observed units for a given time period so that the spatial dependence among these units can be accommodated. This stacking scheme is similar to Wansbeek and Kapteyn (1989) for their unbalanced panel data error component model. Then, the matrix exponential unbalanced panel data model that has the spatial dependence in the dependent variable and in the error terms can be written as

$$e^{\alpha_0 W_t} y_t = X_t \beta_0 + C_t \mu_0 + \lambda_{t0} l_{n_t} + u_t, \quad e^{\tau_0 M_t} u_t = \epsilon_t, \quad t = 1, 2, \dots, T.$$
(2.1)

Here,  $y_t = (y_{1t}, \ldots, y_{n_t t})'$  is the  $n_t \times 1$  dependent variable and  $n_t$  is the number of spatial units available at time t,  $W_t$  and  $M_t$  are  $n_t \times n_t$  spatial weights matrices that have zero diagonal elements,  $X_t$  is the  $n_t \times k$  matrix of exogenous variables with the matching parameter vector  $\beta_0$ ,  $u_t = (u_{1t}, \ldots, u_{n_t t})'$  is the  $n_t \times 1$  vector of regression error terms, and  $\epsilon_t = (\epsilon_{1t}, \ldots, \epsilon_{n_t t})'$  is the  $n_t \times 1$ vector of idiosyncratic error terms (or innovations). In a given time period, we assume that there can be at most n spatial entities, and we use the  $n \times 1$  vector  $\mu_0$  to denote the entity fixed effects. Thus,  $C_t \mu_0$  denotes the entity fixed effects for the spatial units observed at time t, where  $C_t$  is the  $n_t \times n$  sub-matrix obtained by deleting the rows corresponding to the missing entities at time t from the  $n \times n$  identity matrix  $I_n$ . The time fixed effect at period t is denoted by  $\lambda_{t0}$ , and  $l_{n_t}$ is the  $n_t \times 1$  vector of ones. The matrix exponential terms  $e^{\alpha_0 W_t}$  and  $e^{\tau_0 M_t}$  in (2.1) are defined as  $e^{\alpha_0 W_t} = \sum_{i=0}^{\infty} (\alpha_0 W_t)^i / i!$  and  $e^{\tau_0 M_t} = \sum_{i=0}^{\infty} (\tau_0 M_t)^i / i!$ , where  $\alpha_0$  and  $\tau_0$  are the scalar spatial parameters. These terms are always invertable, and the respective inverses are given by  $e^{-\alpha_0 W_t}$ and  $e^{-\tau_0 M_t}$ . Thus, the reduced form of (2.1) is  $y_t = e^{-\alpha_0 W_t} (X_t \beta_0 + C_t \mu_0 + \lambda_{t_0} l_{n_t} + e^{-\tau_0 M_t} \epsilon_t)$  and imposes no restrictions on the parameter space of the spatial parameters.

Let  $blkdiag(A_1, \ldots, A_n)$  be the block-diagonal matrix formed from the set of matrices  $\{A_1, \ldots, A_n\}$ . Define  $\mathbf{e}^{\boldsymbol{\alpha}_0 \mathbf{W}} = blkdiag(e^{\alpha_0 W_1}, \ldots, e^{\alpha_0 W_T}), \mathbf{e}^{\boldsymbol{\tau}_0 \mathbf{M}} = blkdiag(e^{\tau_0 M_1}, \ldots, e^{\tau_0 M_T}), C_{\boldsymbol{\mu}} = (C'_1, \ldots, C'_T)',$ and  $C_{\boldsymbol{\lambda}} = blkdiag(l_{n_1}, \ldots, l_{n_T})$ , Then, stacking the model across t, we obtain

$$\mathbf{e}^{\boldsymbol{\alpha}_0 \mathbf{W}} y = X\beta_0 + C_{\mu}\mu_0 + C_{\lambda}\lambda_0 + u, \quad \mathbf{e}^{\boldsymbol{\tau}_0 \mathbf{M}} u = \epsilon,$$
(2.2)

where  $y = (y'_1, \ldots, y'_T)'$ ,  $X = (X'_1, \ldots, X'_T)'$ ,  $\lambda_0 = (\lambda_{10}, \ldots, \lambda_{T0})'$ ,  $u = (u'_1, \ldots, u'_T)'$ , and  $\epsilon = (\epsilon'_1, \ldots, \epsilon'_T)'$ . For the identification of entity and time fixed effects, we assume that the model is subject to the normalization constraint  $l'_T \lambda_0 = 0$ . Under this normalization constraint, (2.2) can be expressed as

$$\mathbf{e}^{\boldsymbol{\alpha}_0 \mathbf{W}} y = X\beta_0 + C_{\mu}\mu_0 + C_{\lambda}^*\lambda_0^* + u, \quad \mathbf{e}^{\boldsymbol{\tau}_0 \mathbf{M}} u = \epsilon,$$
(2.3)

where  $C_{\lambda}^{*} = \left(-l_{T-1}l_{n_{1}}', \text{blkdiag}'(l_{n_{2}}, \ldots, l_{n_{T}})\right)'$ , and  $\lambda_{0}^{*} = (\lambda_{20}^{*}, \ldots, \lambda_{T0}^{*})'$  is the new vector of time fixed effects. In our analysis, we use  $\theta_{0} = (\beta_{0}', \sigma_{\epsilon_{0}}^{2}, \alpha_{0}, \tau_{0})'$  and  $\delta_{0} = (\mu_{0}', \lambda_{0}^{*'})'$  to denote the true parameter values, and  $\theta = (\beta', \sigma_{\epsilon}^{2}, \alpha, \tau)'$  and  $\delta = (\mu', \lambda^{*'})'$  to denote arbitrary parameter values.

## 3 Estimation under homoskedasticity

In this section, we consider the quasi maximum likelihood estimation of (2.3) under the following assumption.

Assumption 1. The  $\{\epsilon_{it}\}$  are independent and identically distributed (i.i.d) across i and t with mean zero and variance  $\sigma_{\epsilon 0}^2$ , and  $\mathbf{E}|\epsilon_{it}|^{4+\varrho} < \infty$  for some  $\varrho > 0$ .

Under this assumption, the elements of the innovation term are i.i.d across *i* and *t*. The moment condition  $E|\epsilon_{it}|^{4+\varrho} < \infty$  is required for the application of the central limit theorem for linear and quadratic forms (Kelejian and Prucha, 2001). Let  $N = \sum_{t=1}^{T} n_t$ ,  $C = (C_{\mu}, C_{\lambda}^*)$ , and  $\zeta = (\alpha, \tau)'$ . Then, under Assumption 1, the quasi log-likelihood function of the model can be expressed as

$$l(\theta,\delta) = -\frac{N}{2}\ln(2\pi\sigma_{\epsilon}^{2}) + \ln\left|\mathbf{e}^{\boldsymbol{\alpha}\mathbf{W}}\right| + \ln\left|\mathbf{e}^{\boldsymbol{\tau}\mathbf{M}}\right| - \frac{1}{2\sigma_{\epsilon}^{2}}\epsilon'(\beta,\zeta,\delta)\epsilon(\beta,\zeta,\delta), \qquad (3.1)$$

where  $\epsilon(\beta, \zeta, \delta) = \mathbf{e}^{\tau \mathbf{M}} \left( \mathbf{e}^{\alpha \mathbf{W}} y - X\beta - C\delta \right)$  and  $|\cdot|$  is the determinant operator. Since the spatial weights matrices have zero diagonal elements, we have  $\ln |\mathbf{e}^{\alpha \mathbf{W}}| = \sum_{t=1}^{T} \ln \left( e^{\alpha \operatorname{tr}(W_t)} \right) = 0$  and  $\ln |\mathbf{e}^{\tau \mathbf{M}}| = \sum_{t=1}^{T} \ln \left( e^{\tau \operatorname{tr}(M_t)} \right) = 0$ , where  $\operatorname{tr}(\cdot)$  denotes the trace operator. Thus, (3.1) becomes free of the Jacobin terms and simplifies to

$$l(\theta,\delta) = -\frac{N}{2}\ln(2\pi\sigma_{\epsilon}^2) - \frac{1}{2\sigma_{\epsilon}^2}\epsilon'(\beta,\zeta,\delta)\epsilon(\beta,\zeta,\delta).$$
(3.2)

Given  $\theta$ ,  $l(\theta, \delta)$  is partially maximized at

$$\widehat{\delta}(\beta,\zeta) = \left(\mathbb{C}'(\tau)\mathbb{C}(\tau)\right)^{-1}\mathbb{C}'(\tau)\mathbf{e}^{\boldsymbol{\tau}\mathbf{M}}\left(\mathbf{e}^{\boldsymbol{\alpha}\mathbf{W}}y - X\beta\right),\tag{3.3}$$

where  $\mathbb{C}(\tau) = \mathbf{e}^{\tau \mathbf{M}} C$ . Substituting  $\widehat{\delta}(\beta, \zeta)$  into  $l(\theta, \delta)$ , we obtain the following concentrated quasi log-likelihood function

$$l^{c}(\theta) = -\frac{N}{2}\ln(2\pi\sigma_{\epsilon}^{2}) - \frac{1}{2\sigma_{\epsilon}^{2}}\tilde{\epsilon}'(\beta,\zeta)\tilde{\epsilon}(\beta,\zeta), \qquad (3.4)$$

where  $\tilde{\epsilon}(\beta,\zeta) = \mathbb{Q}_{\mathbb{C}}(\tau) \mathbf{e}^{\boldsymbol{\tau}\mathbf{M}} \left( \mathbf{e}^{\boldsymbol{\alpha}\mathbf{W}} y - X\beta \right)$  and  $\mathbb{Q}_{\mathbb{C}}(\tau) = I_N - \mathbb{P}_{\mathbb{C}}(\tau)$  with  $\mathbb{P}_{\mathbb{C}}(\tau) = \mathbb{C}(\tau) \left( \mathbb{C}'(\tau)\mathbb{C}(\tau) \right)^{-1} \mathbb{C}'(\tau)$ . Then, the QMLE  $\hat{\theta}$  of  $\theta$  is defined by  $\hat{\theta} = \operatorname{argmax}_{\theta} l^c(\theta)$ . Let  $S_a^c(\theta) = \frac{\partial l^c(\theta)}{\partial a}$  for  $a \in \{\beta, \sigma_{\epsilon}^2, \alpha, \tau\}$  and  $S^{c}(\theta) = \left(S^{c'}_{\beta}(\theta), S^{c}_{\sigma_{\epsilon}^{2}}(\theta), S^{c}_{\alpha}(\theta), S^{c}_{\tau}(\theta)\right)'.$  Then, using (3.4), we obtain the following score functions:

$$S^{c}(\theta) = \begin{cases} \beta : \frac{1}{\sigma_{\epsilon}^{2}} X' \mathbf{e}^{\tau \mathbf{M}'} \tilde{\epsilon}(\beta, \zeta), \\ \sigma_{\epsilon}^{2} : -\frac{N}{2\sigma_{\epsilon}^{2}} + \frac{1}{2\sigma_{\epsilon}^{4}} \tilde{\epsilon}'(\beta, \zeta) \tilde{\epsilon}(\beta, \zeta), \\ \alpha : -\frac{1}{\sigma_{\epsilon}^{2}} y' \mathbf{e}^{\alpha \mathbf{W}'} \mathbf{W}' \mathbf{e}^{\tau \mathbf{M}'} \tilde{\epsilon}(\beta, \zeta), \\ \tau : -\frac{1}{\sigma_{\epsilon}^{2}} \tilde{\epsilon}'(\beta, \zeta) \mathbf{M} \tilde{\epsilon}(\beta, \zeta), \end{cases}$$
(3.5)

where  $\mathbf{W} = \text{blkdiag}(W_1, \dots, W_T)$  and  $\mathbf{M} = \text{blkdiag}(M_1, \dots, M_T)$ . A necessary condition required for the consistency of the QMLE  $\hat{\theta}$  is  $\text{plim}_{N \to \infty} \frac{1}{N} S^c(\theta_0) = 0$ . To investigate this necessary condition, we derive  $E(S^c(\theta_0))$  as

$$E(S^{c}(\theta_{0})) = \begin{cases} \beta: & 0_{k \times 1}, \\ \sigma_{\epsilon}^{2}: & -\frac{n+T-1}{2\sigma_{\epsilon_{0}}^{2}}, \\ \alpha: & -\operatorname{tr}\left(\mathbb{Q}_{\mathbb{C}}(\tau_{0})\mathbf{e}^{\boldsymbol{\tau}_{0}\mathbf{M}}\mathbf{W}\mathbf{e}^{-\boldsymbol{\tau}_{0}\mathbf{M}}\right), \\ \tau: & -\operatorname{tr}\left(\mathbb{Q}_{\mathbb{C}}(\tau_{0})\mathbf{M}\right). \end{cases}$$
(3.6)

The order of the elements of  $E(S^c(\theta_0))$  is required for the large sample analysis. As such, we make the following assumptions.

**Assumption 2.** (i) n is large, and T can be large or small, (ii)  $n_t$  increases with n at the same rate for all t, and (iii) all spatial entities are observed for at least two time periods.

Assumption 3. (i) The elements of the time varying spatial weights matrices  $\{W_t\}$  and  $\{M_t\}$  are at most of order  $h_n^{-1}$  uniformly such that  $h_n/n \to 0$  as  $n \to \infty$ . (ii) The spatial weights matrices  $\{W_t\}$  and  $\{M_t\}$  are uniformly bounded in both row sum and column sum matrix norms.

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

Assumption 5. (i) Let  $\gamma_{\max}(A)$  and  $\gamma_{\min}(A)$  be respectively the largest and smallest eigenvalues of the symmetric matrix A. There exist some real numbers  $\underline{d}_{\alpha}$ ,  $\underline{d}_{\tau}$ ,  $\overline{d}_{\alpha}$  and  $\overline{d}_{\tau}$  such that  $0 < \underline{d}_{\alpha} \leq$  $\inf \gamma_{\min} \left( \mathbf{e}^{\alpha \mathbf{W}'} \mathbf{e}^{\alpha \mathbf{W}} \right) \leq \sup \gamma_{\max} \left( \mathbf{e}^{\alpha \mathbf{W}'} \mathbf{e}^{\alpha \mathbf{W}} \right) \leq \overline{d}_{\alpha} < \infty$  and  $0 < \underline{d}_{\tau} \leq \inf \gamma_{\min} \left( \mathbf{e}^{\tau \mathbf{M}'} \mathbf{e}^{\tau \mathbf{M}} \right) \leq$  $\sup \gamma_{\max} \left( \mathbf{e}^{\tau \mathbf{M}'} \mathbf{e}^{\tau \mathbf{M}} \right) \leq \overline{d}_{\tau} < \infty$ . (ii)  $e^{\tau M_s} C_s \left( \frac{1}{T} \sum_{t=1}^T C'_t e^{\tau M'_t} K_t(\tau) e^{\tau M_t} C_t \right)^{-1} C'_t e^{\tau M'_t}$  is bounded in both row sum and column sum matrix norms, uniformly in  $\tau \in \Delta_{\tau}$  for all s and t, where  $K_t(\tau) = I_{n_1}$  for t = 1, and  $K_t(\tau) = I_{n_t} - e^{\tau M_t} l_{n_t} \left( l'_{n_t} e^{\tau M'_t} e^{\tau M_t} l_{n_t} \right)^{-1} l'_{n_t} e^{\tau M'_t}$  for  $t = 2, \ldots, T$ .

Assumption 2 lays out the asymptotic setting. The first part requires that n is large and T can be large or small, and does not impose any conditions on the relative growth rates of n and T. The second part requires that  $n_t$  increases with n ensuring that the number of observed spatial units in each period is not small relative to n. The third part ensures that the spatial structure is complete after the entity fixed effects are concentrated out from the model.

Assumption 3 provides the essential properties of the spatial weights matrices. The first part is considered in Lee (2004), and requires that the elements of the spatial weights matrices have order  $O(h_n^{-1})$  uniformly. Here,  $h_n$  can be bounded or divergent with the property  $h_n/n \to 0$  as  $n \to \infty$ . The second part ensures that the spatial correlation is limited to a manageable degree so that the large sample analysis remains tractable (Kelejian and Prucha, 2001). The assumption rules out the cases where as n tends to infinity each unit remains spatially correlated with all remaining (n-1) units and  $h_n$  is divergent. For instance, take  $h_n = O(n)$  and each unit remains spatially correlated with all remaining (n-1) units. Although Assumption 3 (ii) is satisfied in this case, Assumption 3 (i) is violated. As another example, take  $h_n = O(n^{1/2})$  and each unit remains spatially correlated with all remaining (n-1) units. Although Assumption 3 (i) is satisfied in this case, Assumption 3 (ii) is violated. Such cases can arise for example when the domain of the spatial units is kept fixed and n tends to infinity (Cressie, 1993, p.350). Assumption 4 requires that the parameter space of the spatial parameters is compact, as in Debarsy et al. (2015). Note that the parameter space does not depend on the features of the spatial weights matrices. This assumption and the second part of Assumption 3 imply that the matrix exponential terms in our model are uniformly bounded in both row sum and column sum matrix norms. This can be seen from  $\left\|e^{\alpha W_t}\right\| = \left\|\sum_{i=0}^{\infty} \alpha^i W_t^i / i!\right\| \le \sum_{i=0}^{\infty} |\alpha|^i \|W_t\|^i / i! = e^{|\alpha| \|W_t\|}$ , which is bounded if  $|\alpha|$  and  $\|W_t\|$ are bounded, where  $\|\cdot\|$  is either the row sum or the column sum matrix norm.

The first part of Assumption 5 states that the minimum and the maximum eigenvalues of  $\mathbf{e}^{\alpha \mathbf{W}'} \mathbf{e}^{\alpha \mathbf{W}}$  and  $\mathbf{e}^{\tau \mathbf{M}'} \mathbf{e}^{\tau \mathbf{M}}$  are bounded uniformly. Note that these terms are positive definite, so their smallest eigenvalues are positive. In the first part of Assumption 5, we require that the eigenvalues of these terms are strictly uniformly positive over the parameter space. In addition,

the largest eigenvalues of  $\mathbf{e}^{\alpha \mathbf{W}'} \mathbf{e}^{\alpha \mathbf{W}}$  and  $\mathbf{e}^{\tau \mathbf{M}'} \mathbf{e}^{\tau \mathbf{M}}$  are bounded since these terms are uniformly bounded in both row and column sums under Assumptions 3 (ii) and 4. We further require that the eigenvalues of these terms are uniformly bounded above over  $\Delta$ . Finally, the last part of Assumption 5 ensures that  $\mathbb{Q}_{\mathbb{C}}(\tau)$  is uniformly bounded in both row sum and column sum matrix norms.

Under these assumptions, it follows that  $\frac{1}{N} \mathbb{E} \left( S_{\sigma_{\epsilon}^{c}}^{c}(\theta_{0}) \right) = O\left((n+T)/N\right) = O(1/T)$ . Also,  $\frac{1}{N} \mathbb{E} \left( S_{\alpha}^{c}(\theta_{0}) \right) = -\frac{1}{N} \mathrm{tr} \left( \mathbb{Q}_{\mathbb{C}}(\tau_{0}) \mathbf{e}^{\tau_{0}\mathbf{M}} \mathbf{W} \mathbf{e}^{-\tau_{0}\mathbf{M}} \right) = -\frac{1}{N} \mathrm{tr} \left( \mathbf{e}^{\tau_{0}\mathbf{M}} \mathbf{W} \mathbf{e}^{-\tau_{0}\mathbf{M}} \right) + \frac{1}{N} \mathrm{tr} \left( \mathbb{P}_{\mathbb{C}}(\tau_{0}) \mathbf{e}^{\tau_{0}\mathbf{M}} \mathbf{W} \mathbf{e}^{-\tau_{0}\mathbf{M}} \right) = \frac{1}{N} \mathrm{tr} \left( \mathbb{P}_{\mathbb{C}}(\tau_{0}) \mathbf{e}^{\tau_{0}\mathbf{M}} \mathbf{W} \mathbf{e}^{-\tau_{0}\mathbf{M}} \right) = O\left(1/\max\{T, h_{n}\}\right)$  by Lemma A.2 (iv) of the web appendix. Similarly,  $\frac{1}{N} \mathbb{E} \left( S_{\tau}^{c}(\theta_{0}) \right) = -\frac{1}{N} \mathrm{tr} \left( \mathbb{Q}_{\mathbb{C}}(\tau_{0}) \mathbf{M} \right) = -\frac{1}{N} \mathrm{tr} \left( \mathbf{M} \right) + \frac{1}{N} \mathrm{tr} \left( \mathbb{P}_{\mathbb{C}}(\tau_{0}) \mathbf{M} \right) = O\left(1/\max\{T, h_{n}\}\right)$  by Lemma A.2 (iv) of the web appendix. Thus, the necessary condition  $\mathrm{plim}_{N\to\infty} \frac{1}{N} S^{c}(\theta_{0}) = 0$  for the consistency of the QMLE is satisfied if T is divergent. Also note that we have  $\frac{1}{N} \mathbb{E} \left( S_{\tau}^{c}(\theta_{0}) \right) = o(1)$  and  $\frac{1}{N} \mathbb{E} \left( S_{\alpha}^{c}(\theta_{0}) \right) = o(1)$  when either T or  $h_{n}$  is unbounded. Therefore, it is important to adjust the score functions such that  $\mathrm{plim}_{N\to\infty} \frac{1}{N} S^{c}(\theta_{0}) = 0$  holds in all cases. To that end, we consider the general form of the adjusted quasi-score functions  $S^{*}(\theta_{0}) = S^{c}(\theta_{0}) - \mathbb{E} \left( S^{c}(\theta_{0}) \right)$ :

$$S^{*}(\theta) = \begin{cases} \beta : \frac{1}{\sigma_{\epsilon}^{2}} X' \mathbf{e}^{\boldsymbol{\tau} \mathbf{M}'} \tilde{\epsilon}(\beta, \zeta), \\ \sigma_{\epsilon}^{2} : -\frac{N_{1}}{2\sigma_{\epsilon}^{2}} + \frac{1}{2\sigma_{\epsilon}^{4}} \tilde{\epsilon}'(\beta, \zeta) \tilde{\epsilon}(\beta, \zeta), \\ \alpha : -\frac{1}{\sigma_{\epsilon}^{2}} y' \mathbf{e}^{\boldsymbol{\alpha} \mathbf{W}'} \mathbf{W}' \mathbf{e}^{\boldsymbol{\tau} \mathbf{M}'} \tilde{\epsilon}(\beta, \zeta) + \operatorname{tr}\left(\mathbb{Q}_{\mathbb{C}}(\tau) \mathbf{e}^{\boldsymbol{\tau} \mathbf{M}} \mathbf{W} \mathbf{e}^{-\boldsymbol{\tau} \mathbf{M}}\right), \\ \tau : -\frac{1}{\sigma_{\epsilon}^{2}} \tilde{\epsilon}'(\beta, \zeta) \mathbf{M} \tilde{\epsilon}(\beta, \zeta) + \operatorname{tr}\left(\mathbb{Q}_{\mathbb{C}}(\tau) \mathbf{M}\right), \end{cases}$$
(3.7)

where  $N_1 = N - n - T + 1$ . From the first two adjusted score functions in  $S^*(\theta)$ , we can formulate the estimators of  $\beta$  and  $\sigma_{\epsilon}^2$  for a given  $\zeta$  value as

$$\widehat{\boldsymbol{\beta}}^{*}(\boldsymbol{\zeta}) = \left(\boldsymbol{X}' \mathbf{e}^{\boldsymbol{\tau} \mathbf{M}'} \mathbb{Q}_{\mathbb{C}}(\boldsymbol{\tau}) \mathbf{e}^{\boldsymbol{\tau} \mathbf{M}} \boldsymbol{X}\right)^{-1} \boldsymbol{X}' \mathbf{e}^{\boldsymbol{\tau} \mathbf{M}'} \mathbb{Q}_{\mathbb{C}}(\boldsymbol{\tau}) \mathbf{e}^{\boldsymbol{\tau} \mathbf{M}} \mathbf{e}^{\boldsymbol{\alpha} \mathbf{W}} \boldsymbol{y},$$
(3.8)

$$\widehat{\sigma}_{\epsilon}^{*2}(\zeta) = \widehat{\epsilon}'(\zeta)\widehat{\epsilon}(\zeta)/N_1, \qquad (3.9)$$

where  $\hat{\epsilon}(\zeta) = \tilde{\epsilon}(\hat{\beta}^*(\zeta), \zeta)$ . Substituting  $\hat{\beta}^*(\zeta)$  and  $\hat{\sigma}_{\epsilon}^{*2}(\zeta)$  into  $\alpha$  and  $\tau$  elements of  $S^*(\theta)$ , we obtain

$$S^{c*}(\zeta) = \begin{cases} \alpha : & -\frac{1}{\widehat{\sigma}_{\epsilon}^{*2}(\zeta)} y' \mathbf{e}^{\alpha \mathbf{W}'} \mathbf{W}' \mathbf{e}^{\tau \mathbf{M}'} \widehat{\epsilon}(\zeta) + \operatorname{tr} \left( \mathbb{Q}_{\mathbb{C}}(\tau) \mathbf{e}^{\tau \mathbf{M}} \mathbf{W} \mathbf{e}^{-\tau \mathbf{M}} \right), \\ \tau : & -\frac{1}{\widehat{\sigma}_{\epsilon}^{*2}(\zeta)} \widehat{\epsilon}'(\zeta) \mathbf{M} \widehat{\epsilon}(\zeta) + \operatorname{tr} \left( \mathbb{Q}_{\mathbb{C}}(\tau) \mathbf{M} \right). \end{cases}$$
(3.10)

Our suggested M-estimator (ME)  $\hat{\zeta}^*$  of  $\zeta_0$  is obtained by solving  $S^{c*}(\zeta) = 0$ , i.e.,  $\hat{\zeta}^* = \operatorname{argsolve} \{S^{c*}(\zeta) = 0\}$ . Using  $\hat{\zeta}^*$ , we can formulate the M-estimators of  $\beta_0$  and  $\sigma_{\epsilon 0}^2$  by  $\hat{\beta}^* = \hat{\beta}^*(\hat{\zeta}^*)$  and  $\hat{\sigma}_{\epsilon}^{*2} = \hat{\sigma}_{\epsilon}^{*2}(\hat{\zeta}^*)$ . To prove the consistency of  $\hat{\zeta}^*$ , we need the population counterpart of  $S^{c*}(\zeta)$ . Let  $\overline{S}^*(\theta) = E(S^*(\theta))$ . The consistency of  $\hat{\zeta}^*$  requires that  $S^{*c}(\zeta)$  converges to  $\overline{S}^{c*}(\zeta)$  uniformly over  $\Delta$ , i.e.,  $\sup_{\zeta \in \Delta} \frac{1}{N_1} \|S^{*c}(\zeta) - \overline{S}^{*c}(\zeta)\| \xrightarrow{p} 0$  (van der Vaart, 1998). Given  $\zeta$ , we can derive the following population counterparts of (3.8) and (3.9) as:

$$\bar{\beta}^{*}(\zeta) = \left(X' \mathbf{e}^{\boldsymbol{\tau} \mathbf{M}'} \mathbb{Q}_{\mathbb{C}}(\tau) \mathbf{e}^{\boldsymbol{\tau} \mathbf{M}} X\right)^{-1} X' \mathbf{e}^{\boldsymbol{\tau} \mathbf{M}'} \mathbb{Q}_{\mathbb{C}}(\tau) \mathbf{G}(\zeta) \mathbf{E}(y), \tag{3.11}$$

$$\bar{\sigma}^{*2}(\zeta) = \mathbf{E}\left(\bar{\epsilon}'(\zeta)\bar{\epsilon}(\zeta)\right) / N_{1} \tag{3.12}$$

$$\overline{\sigma}_{\epsilon}^{*2}(\zeta) = \mathbf{E}\left(\overline{\epsilon}'(\zeta)\overline{\epsilon}(\zeta)\right)/N_1,\tag{3.12}$$

where  $\mathbf{G}(\zeta) = \mathbf{e}^{\boldsymbol{\tau}\mathbf{M}}\mathbf{e}^{\boldsymbol{\alpha}\mathbf{W}}$  and  $\bar{\epsilon}(\zeta) = \tilde{\epsilon}(\bar{\beta}^*(\zeta), \zeta)$ . Note that we can express  $\bar{\epsilon}(\zeta)$  as

$$\bar{\epsilon}(\zeta) = \mathbb{Q}_{\mathbb{C}}(\tau) \mathbf{e}^{\tau \mathbf{M}} \left( \mathbf{e}^{\alpha \mathbf{W}} y - X \bar{\beta}^{*}(\zeta) \right) = \mathbb{Q}_{\mathbb{C}}(\tau) \mathbf{G}(\zeta) y - \mathbb{P}_{\mathbb{X}}(\tau) \mathbb{Q}_{\mathbb{C}}(\tau) \mathbf{G}(\zeta) \mathbf{E}(y) 
= \left( \mathbb{P}_{\mathbb{X}}(\tau) + \mathbb{Q}_{\mathbb{X}}(\tau) \right) \mathbb{Q}_{\mathbb{C}}(\tau) \mathbf{G}(\zeta) y - \mathbb{P}_{\mathbb{X}}(\tau) \mathbb{Q}_{\mathbb{C}}(\tau) \mathbf{G}(\zeta) \mathbf{E}(y) 
= \mathbb{P}_{\mathbb{X}}(\tau) \mathbb{Q}_{\mathbb{C}}(\tau) \mathbf{G}(\zeta) \left( y - \mathbf{E}(y) \right) + \mathbb{Q}_{\mathbb{X}}(\tau) \mathbb{Q}_{\mathbb{C}}(\tau) \mathbf{G}(\zeta) y,$$
(3.13)

where  $\mathbb{X}(\tau) = \mathbb{Q}_{\mathbb{C}}(\tau) \mathbf{e}^{\tau \mathbf{M}} X$ ,  $\mathbb{Q}_{\mathbb{X}}(\tau) = I_N - \mathbb{P}_{\mathbb{X}}(\tau)$ , and  $\mathbb{P}_{\mathbb{X}}(\tau)$  is the projection matrix based on  $\mathbb{X}(\tau)$ . Substituting  $\overline{\beta}^*(\zeta)$  and  $\overline{\sigma}_{\epsilon}^{*2}(\zeta)$  into the  $\alpha$  and  $\tau$  elements of  $\overline{S}^*(\theta)$ , we get the population counterpart of  $S^{c*}(\zeta)$ :

$$\bar{S}^{c*}(\zeta) = \begin{cases} \alpha : & -\frac{1}{\bar{\sigma}_{\epsilon}^{*2}(\zeta)} \mathbf{E} \left( y' \mathbf{e}^{\alpha \mathbf{W}'} \mathbf{W}' \mathbf{e}^{\tau \mathbf{M}'} \bar{\epsilon}(\zeta) \right) + \mathrm{tr} \left( \mathbb{Q}_{\mathbb{C}}(\tau) \mathbf{e}^{\tau \mathbf{M}} \mathbf{W} \mathbf{e}^{-\tau \mathbf{M}} \right), \\ \tau : & -\frac{1}{\bar{\sigma}_{\epsilon}^{*2}(\zeta)} \mathbf{E} \left( \bar{\epsilon}'(\zeta) \mathbf{M} \bar{\epsilon}(\zeta) \right) + \mathrm{tr} \left( \mathbb{Q}_{\mathbb{C}}(\tau) \mathbf{M} \right). \end{cases}$$
(3.14)

Using the result in (3.14), we state the identification condition in our setting by the following assumption.

**Assumption 6.**  $\inf_{\zeta: d(\zeta, \zeta_0) \ge \vartheta} \|\overline{S}^{*c}(\zeta)\| > 0$  for every  $\vartheta > 0$ , where  $d(\zeta, \zeta_0)$  is a measure of distance between  $\zeta$  and  $\zeta_0$ .

Assumption 6 is a high level assumption and can be difficult to verify in practice. However, we can investigate sufficient low level conditions from the requirement that  $\overline{S}^{*c}(\zeta) \neq 0$  whenever  $\zeta \neq \zeta_0$ . In Section F of the accompanying web appendix, we provide such conditions under which Assumption 6 holds.

Using (3.10) and (3.14), we can express  $S^{*c}(\zeta) - \overline{S}^{*c}(\zeta)$  as

$$S^{*c}(\zeta) - \overline{S}^{*c}(\zeta) = \begin{cases} \alpha : & -\frac{1}{\widehat{\sigma}_{\epsilon}^{*2}(\zeta)} y' \mathbf{e}^{\alpha \mathbf{W}'} \mathbf{W}' \mathbf{e}^{\tau \mathbf{M}'} \widehat{\epsilon}(\zeta) + \frac{1}{\overline{\sigma}_{\epsilon}^{*2}(\zeta)} \mathbf{E} \left( y' \mathbf{e}^{\alpha \mathbf{W}'} \mathbf{W}' \mathbf{e}^{\tau \mathbf{M}'} \overline{\epsilon}(\zeta) \right), \\ \tau : & -\frac{1}{\widehat{\sigma}_{\epsilon}^{*2}(\zeta)} \widehat{\epsilon}'(\zeta) \mathbf{M} \widehat{\epsilon}(\zeta) + \frac{1}{\overline{\sigma}_{\epsilon}^{*2}(\zeta)} \overline{\epsilon}'(\zeta) \mathbf{M} \overline{\epsilon}(\zeta). \end{cases}$$
(3.15)

The results in (3.15) coupled with (3.13) will be useful for showing the consistency of  $\hat{\zeta}^*$ . To that end, we also require the following assumption.

Assumption 7. X is exogenous, with uniformly bounded elements, and has full column rank. Also  $\lim_{N\to\infty} \frac{1}{N} \mathbb{X}'(\tau) \mathbb{X}(\tau) \text{ exists and is nonsingular, uniformly in } \tau \in \Delta_{\tau}.$ 

Assumption 7 provides a regularity condition for the exogenous variables (Debarsy et al., 2015). In particular this assumption ensures that  $\mathbb{Q}_{\mathbb{X}}(\tau)$  is uniformly bounded in row sum and column sum matrix norms, uniformly in  $\tau \in \Delta_{\tau}$ . Under our stated assumptions, the uniform convergence  $\sup_{\zeta \in \Delta} \frac{1}{N_1} \|S^{*c}(\zeta) - \overline{S}^{*c}(\zeta)\| \xrightarrow{p} 0$  ensures the consistency of  $\widehat{\zeta}^*$ . The following theorem formalizes the consistency of  $\widehat{\theta}^*$  under our stated assumptions.

**Theorem 3.1.** Under Assumptions 1–7, it follows that  $\widehat{\theta}^* \xrightarrow{p} \theta_0$  as  $N \to \infty$ .

*Proof.* See Section C in the web appendix.

To derive the asymptotic distribution of  $\hat{\theta}^*$ , we apply the mean value theorem to  $S^*(\hat{\theta}^*) = 0$  at  $\theta_0$ , to obtain  $\sqrt{N_1}(\hat{\theta}^* - \theta_0) = -\left(\frac{1}{N_1}\frac{\partial S^*(\bar{\theta})}{\partial \theta'}\right)^{-1}\frac{1}{\sqrt{N_1}}S^*(\theta_0)$ , where  $\bar{\theta}$  lies between  $\theta_0$  and  $\hat{\theta}^*$  elementwise (Jennrich, 1969, Lemma 3). Let  $\phi = X\beta_0 + C\delta_0$ . Then, y can be expressed as  $y = \mathbf{e}^{-\alpha_0}\mathbf{W}(\phi + \mathbf{e}^{-\tau_0}\mathbf{M}\epsilon)$ . Substituting  $y = \mathbf{e}^{-\alpha_0}\mathbf{W}(\phi + \mathbf{e}^{-\tau_0}\mathbf{M}\epsilon)$  and  $\tilde{\epsilon}(\beta_0, \zeta_0) = \mathbb{Q}_{\mathbb{C}}(\tau_0)\epsilon$  into  $S^*(\theta_0)$ , we can express  $S^*(\theta_0)$  in terms of linear and quadratic forms of  $\epsilon$  in the following way:

$$S^{*}(\theta_{0}) = \begin{cases} \beta : \frac{1}{\sigma_{\epsilon_{0}}^{2}} \mathbb{X}'(\tau_{0})\epsilon, \\ \sigma_{\epsilon}^{2} : -\frac{N_{1}}{2\sigma_{\epsilon_{0}}^{2}} + \frac{1}{2\sigma_{\epsilon_{0}}^{4}}\epsilon' \mathbb{Q}_{\mathbb{C}}(\tau_{0})\epsilon, \\ \alpha : -\frac{1}{\sigma_{\epsilon_{0}}^{2}}\epsilon' \mathbf{S}'(\tau_{0})\mathbb{Q}_{\mathbb{C}}(\tau_{0})\epsilon - \frac{1}{\sigma_{\epsilon_{0}}^{2}}\phi' \mathbf{e}^{\tau_{0}\mathbf{M}'} \mathbf{S}'(\tau_{0})\mathbb{Q}_{\mathbb{C}}(\tau_{0})\epsilon + \operatorname{tr}\left(\mathbb{Q}_{\mathbb{C}}(\tau_{0})\mathbf{S}(\tau_{0})\right), \\ \tau : -\frac{1}{\sigma_{\epsilon_{0}}^{2}}\epsilon' \mathbb{Q}'_{\mathbb{C}}(\tau_{0})\mathbf{M}\mathbb{Q}_{\mathbb{C}}(\tau_{0})\epsilon + \operatorname{tr}\left(\mathbb{Q}_{\mathbb{C}}(\tau_{0})\mathbf{M}\right), \end{cases}$$
(3.16)

where  $\mathbf{S}(\tau_0) = \mathbf{e}^{\boldsymbol{\tau}_0 \mathbf{M}} \mathbf{W} \mathbf{e}^{-\boldsymbol{\tau}_0 \mathbf{M}}$ . Then, we can apply the central limit theorem (CLT) for linearquadratic forms in Kelejian and Prucha (2001) to establish the asymptotic normality of  $\frac{1}{\sqrt{N_1}} S^*(\theta_0)$ . Also, our assumptions ensure that  $\frac{1}{N_1} \frac{\partial S^*(\bar{\theta})}{\partial \theta'} - \frac{1}{N_1} \mathbf{E} \left( \frac{\partial S^*(\theta_0)}{\partial \theta'} \right) = o_p(1)$ . These results, as shown in the proof, lead to the following theorem.

**Theorem 3.2.** Under Assumptions 1–7, as  $N \to \infty$ , we have

$$\sqrt{N_1}(\widehat{\theta}^* - \theta_0) \xrightarrow{d} N\left(0, \lim_{N \to \infty} \Psi^{*-1}(\theta_0)\Omega^*(\theta_0)\Psi^{*'-1}(\theta_0)\right), \qquad (3.17)$$

where  $\Psi^*(\theta_0) = -\frac{1}{N_1} \operatorname{E}\left(\frac{\partial S^*(\theta_0)}{\partial \theta'}\right)$  and  $\Omega^*(\theta_0) = \frac{1}{N_1} \operatorname{Var}\left(S^*(\theta_0)\right)$  are assumed to exist and  $\Psi^*(\theta_0)$  is assumed to be positive definite for sufficiently large N.

*Proof.* See Section C in the web appendix.

Theorem 3.2 indicates that we need a consistent estimate of the variance-covariance matrix  $\Psi^{*-1}(\theta_0)\Omega^*(\theta_0)\Psi^{*'-1}(\theta_0)$  for inference. In the case of  $\Psi^*(\theta_0)$ , we can use the observed counterpart given by  $\Psi^*(\widehat{\theta}^*) = -\frac{1}{N_1}\frac{\partial S^*(\theta)}{\partial \theta'}|_{\theta=\widehat{\theta}^*}$ . Let  $H^*_a(\theta) = -\frac{1}{N_1}\frac{\partial S^*(\theta)}{\partial a'}$  for  $a \in \{\beta, \sigma_{\epsilon}^2, \alpha, \tau\}$  and  $A^s = A + A'$  for any square matrix A. Then, it is easy to show that

$$\begin{split} &N_{1}H_{\beta\beta}^{*}(\theta) = \frac{1}{\sigma_{\epsilon}^{2}}\mathbb{X}^{'}(\tau)\mathbb{X}(\tau), \quad N_{1}H_{\beta\sigma_{\epsilon}^{2}}^{*}(\theta) = N_{1}H_{\sigma_{\epsilon}^{2}\beta}^{*'}(\theta) = \frac{1}{\sigma_{\epsilon}^{4}}\mathbb{X}^{'}(\tau)\tilde{\epsilon}(\beta,\zeta), \\ &N_{1}H_{\beta\alpha}^{*}(\theta) = H_{\alpha\beta}^{*'}(\theta) = -\frac{1}{\sigma_{\epsilon}^{2}}\mathbb{X}^{'}(\tau)y(\zeta), \quad N_{1}H_{\beta\tau}^{*}(\theta) = H_{\tau\beta}^{*'}(\theta) = -\frac{1}{\sigma_{\epsilon}^{2}}\mathbb{X}^{'}(\tau)\mathbf{M}^{s}\tilde{\epsilon}(\beta,\zeta), \\ &N_{1}H_{\sigma_{\epsilon}^{2}\sigma_{\epsilon}^{2}}^{*}(\theta) = -\frac{N_{1}}{2\sigma_{\epsilon}^{4}} + \frac{1}{\sigma_{\epsilon}^{6}}\tilde{\epsilon}^{'}(\beta,\zeta)\tilde{\epsilon}(\beta,\zeta), \quad N_{1}H_{\sigma_{\epsilon}^{2}\alpha}^{*}(\theta) = N_{1}H_{\alpha\sigma_{\epsilon}^{2}}^{*}(\theta) = -\frac{1}{\sigma_{\epsilon}^{4}}y^{'}(\zeta)\tilde{\epsilon}(\beta,\zeta), \\ &N_{1}H_{\sigma\epsilon^{2}\tau}^{*}(\theta) = N_{1}H_{\tau\sigma_{\epsilon}^{2}}^{*}(\theta) = -\frac{1}{2\sigma_{\epsilon}^{4}}\tilde{\epsilon}^{'}(\beta,\zeta)\mathbf{M}^{s}\tilde{\epsilon}(\beta,\zeta), \quad N_{1}H_{\alpha\alpha}^{*}(\theta) = \frac{1}{\sigma_{\epsilon}^{2}}[y_{2}^{'}(\zeta)\tilde{\epsilon}(\beta,\zeta) + y^{'}(\zeta)y(\zeta)] \\ &N_{1}H_{\alpha\tau}^{*}(\theta) = \frac{1}{\sigma_{\epsilon}^{2}}y^{'}(\zeta)\mathbf{M}^{s}\tilde{\epsilon}(\beta,\zeta) + \operatorname{tr}\left(\mathbb{P}_{\mathbb{C}}(\tau)\mathbf{M}^{s}\mathbb{Q}_{\mathbb{C}}(\tau)\mathbf{S}(\tau)\right), \\ &N_{1}H_{\tau\alpha}^{*}(\theta) = \frac{1}{\sigma_{\epsilon}^{2}}y^{'}(\zeta)\mathbf{M}^{s}\tilde{\epsilon}(\beta,\zeta), \\ &N_{1}H_{\tau\alpha}^{*}(\theta) = \frac{1}{\sigma_{\epsilon}^{2}}y^{'}(\zeta)\mathbf{M}^{s}\tilde{\epsilon}(\beta,\zeta), \\ &N_{1}H_{\tau\alpha}^{*}(\theta) = \frac{1}{\sigma_{\epsilon}^{2}}y^{'}(\zeta)\mathbf{M}^{s}\tilde{\epsilon}(\beta,\zeta), \\ &N_{1}H_{\tau\alpha}^{*}(\theta) = \frac{1}{\sigma_{\epsilon}^{2}}}\tilde{\epsilon}^{'}(\beta,\zeta)\mathbb{D}^{'}(\tau)\mathbf{M}^{s}\tilde{\epsilon}(\beta,\zeta) + \operatorname{tr}\left(\mathbb{Q}_{\mathbb{C}}(\tau)\mathbf{M}\mathbb{P}_{\mathbb{C}}(\tau)\mathbf{M}^{s}\right), \end{split}$$

where  $y(\zeta) = \mathbb{Q}_{\mathbb{C}}(\tau) \mathbf{e}^{\boldsymbol{\tau}\mathbf{M}} \mathbf{W} \mathbf{e}^{\boldsymbol{\alpha}\mathbf{W}} y, y_2(\zeta) = \mathbb{Q}_{\mathbb{C}}(\tau) \mathbf{e}^{\boldsymbol{\tau}\mathbf{M}} \mathbf{W}^2 \mathbf{e}^{\boldsymbol{\alpha}\mathbf{W}} y, \mathbb{D}(\tau) = \mathbb{Q}_{\mathbb{C}}(\tau) \mathbf{M} - \mathbb{P}_{\mathbb{C}}(\tau) \mathbf{M}'$ , and  $\mathbf{R}(\tau) = \mathbf{e}^{\boldsymbol{\tau}\mathbf{M}} \mathbf{M} \mathbf{e}^{-\boldsymbol{\tau}\mathbf{M}}$ . Under our stated assumptions, it can be shown that  $\Psi^*(\widehat{\theta}^*) = \Psi^*(\theta_0) + o_p(1)$  (see the proof of Theorem 3.2 for the details).

In the case of  $\Omega^*(\theta_0)$ , we suggest a plug-in estimator. Let  $\operatorname{vec}_D(A)$  be the vector that contains the diagonal elements of a square matrix A. Define  $q_1 = \operatorname{vec}_D(\mathbb{Q}_{\mathbb{C}}(\tau_0))$ ,  $\mathcal{Q}_2(\tau_0) = \mathbb{Q}_{\mathbb{C}}(\tau_0)\mathbf{S}(\tau_0)$ ,  $\mathcal{Q}_3(\tau_0) = \mathbb{Q}_{\mathbb{C}}(\tau_0)\mathbf{e}^{\tau_0\mathbf{M}}\mathbf{W}$ ,  $\mathcal{Q}_4(\tau_0) = \mathbb{Q}_{\mathbb{C}}(\tau_0)\mathbf{M}\mathbb{Q}_{\mathbb{C}}(\tau_0)$  and  $q_r = \operatorname{vec}_D(\mathcal{Q}_r(\tau_0))$  for r = 2, 3, 4. Let  $\rho = \operatorname{E}(\epsilon_i^3)/\sigma_{\epsilon_0}^3$  and  $\kappa = \operatorname{E}(\epsilon_i^4)/\sigma_{\epsilon_0}^4 - 3$  be the skewness and excess kurtosis parameters, respectively. Let  $\Omega_{ab}^*(\theta_0) = \frac{1}{N_1}\operatorname{E}\left(S_a^*(\theta_0)S_b^{*'}(\theta_0)\right)$  for  $a, b \in \{\beta, \sigma_{\epsilon}^2, \alpha, \tau\}$ . Then, using Lemma A.6 of the web appendix, we can derive the following closed-form expressions for the elements of  $\Omega^*(\theta_0)$ :

$$N_{1}\Omega_{\beta\beta}^{*}(\theta_{0}) = \frac{1}{\sigma_{\epsilon0}^{2}}\mathbb{X}'(\tau_{0})\mathbb{X}(\tau_{0}), \quad N_{1}\Omega_{\beta\sigma\epsilon}^{*}(\theta_{0}) = \frac{\rho}{2\sigma_{\epsilon0}^{3}}\mathbb{X}'(\tau_{0})q_{1},$$

$$N_{1}\Omega_{\beta\alpha}^{*}(\theta_{0}) = -\frac{1}{\sigma_{\epsilon0}^{2}}\mathbb{X}'(\tau_{0})Q_{3}(\tau_{0})\phi - \frac{\rho}{\sigma_{\epsilon0}}\mathbb{X}'(\tau_{0})q_{2},$$

$$N_{1}\Omega_{\beta\tau}^{*}(\theta_{0}) = -\frac{\rho}{\sigma_{\epsilon0}}\mathbb{X}'(\tau_{0})q_{4}, \quad N_{1}\Omega_{\sigma\epsilon}^{*}\sigma\epsilon^{*}(\theta_{0}) = \frac{1}{4\sigma_{\epsilon0}^{4}}(\kappa q_{1}'q_{1} + 2N_{1}),$$

$$N_{1}\Omega_{\sigma\epsilon}^{*}(\theta_{0}) = -\frac{\rho}{2\sigma_{\epsilon0}^{3}}q_{1}'Q_{3}(\tau_{0})\phi - \frac{1}{2\sigma_{\epsilon0}^{2}}\left(\kappa q_{1}'q_{2} + 2\operatorname{tr}(Q_{2}(\tau_{0}))\right),$$

$$N_{1}\Omega_{\sigma\epsilon}^{*}(\theta_{0}) = -\frac{1}{2\sigma_{\epsilon0}^{2}}\left(\kappa q_{1}'q_{4} + 2\operatorname{tr}(\mathbb{Q}_{\mathbb{C}}(\tau_{0})\mathbf{M})\right),$$

$$N_{1}\Omega_{\alpha\alpha}^{*}(\theta_{0}) = \frac{1}{\sigma_{\epsilon0}^{2}}\phi'Q_{3}'(\tau_{0})Q_{3}(\tau_{0})\phi + \frac{2\rho}{\sigma_{\epsilon0}}q_{2}'Q_{3}(\tau_{0})\phi + \kappa q_{2}'q_{2} + \operatorname{tr}(Q_{2}(\tau_{0})Q_{2}^{*}(\tau_{0})),$$

$$N_{1}\Omega_{\alpha\tau}^{*}(\theta_{0}) = \frac{\rho}{\sigma_{\epsilon0}}q_{4}'Q_{3}(\tau_{0})\phi + \kappa q_{2}'q_{4} + \operatorname{tr}(Q_{2}(\tau_{0})Q_{4}^{*}(\tau_{0})),$$

$$N_{1}\Omega_{\tau\tau}^{*}(\theta_{0}) = \kappa q_{4}'q_{4} + \operatorname{tr}(Q_{4}(\tau_{0})Q_{4}^{*}(\tau_{0})).$$
(3.18)

Since  $\phi$  appears in the elements of  $\Omega^*(\theta_0)$ , we need a consistent estimator of  $\phi$  to derive a consistent estimator of  $\Omega^*(\theta_0)$ . Let  $\hat{\delta}^* = \hat{\delta}(\hat{\beta}^*, \hat{\zeta}^*)$  be the plug-in estimator of  $\delta_0$  obtained from (3.3), and  $\hat{\rho}$  and  $\hat{\kappa}$  be consistent estimators of  $\rho$  and  $\kappa$ , respectively. Then, the plug-in estimator of  $\Omega^*(\theta_0)$  is given by  $\Omega^*(\hat{\theta}^*) = \Omega^*(\theta)|_{\theta = \hat{\theta}^*, \delta = \hat{\delta}^*, \rho = \hat{\rho}, \kappa = \hat{\kappa}}$ . In the next theorem, we show that this plug-in estimator is a biased estimator of  $\Omega^*(\theta_0)$  under our stated assumptions.

**Theorem 3.3.** Under Assumptions 1–7, as  $N \to \infty$ , we have

$$\Omega^*(\widehat{\theta}^*) = \Omega^*(\theta_0) + \operatorname{Bias}^*(\tau_0) + o_p(1), \tag{3.19}$$

where  $\operatorname{Bias}^*(\tau_0)$  is an  $(k+3) \times (k+3)$  matrix with zero entries everywhere except the  $(\alpha, \alpha)$  entry, which is  $\frac{1}{N_1} \operatorname{tr} \left( \mathbb{P}_{\mathbb{C}}(\tau_0) \mathcal{Q}'_2(\tau_0) \mathcal{Q}_2(\tau_0) \right).$ 

*Proof.* See Section C in the web appendix.

The bias term arises because we may not be able to consistently estimate  $\delta_0$  when T is fixed. Under our assumptions, both  $\mathcal{Q}_2(\tau_0)$  and  $\mathbb{P}_{\mathbb{C}}(\tau_0)$  are bounded in row sum and column sum matrix norms. Moreover, in Lemma A.2 (iii) of the web appendix, we show that the elements of  $\mathbb{P}_{\mathbb{C}}(\tau_0)$ are  $O(\max\{1/n, 1/T\})$  uniformly. Also, it follows from Lemma A.1 of the web appendix that the elements of  $\mathcal{Q}_2(\tau_0)$  are  $O(1/h_n)$  uniformly. Hence,  $\frac{1}{N_1} \operatorname{tr} \left( \mathbb{P}_{\mathbb{C}}(\tau_0) \mathcal{Q}_2(\tau_0) \mathcal{Q}_2(\tau_0) \right)$  is  $O(1/\max\{T, h_n\})$ by Lemma A.2 (iv) of the web appendix. Therefore, in settings with large T and bounded  $h_n$ , or with fixed T and divergent  $h_n$ , the bias term becomes negligible. However, in settings with fixed Tand bounded  $h_n$ , the bias correction is necessary for valid inference.

The bias corrected estimator is given by  $\widehat{\Omega}^* = \Omega^*(\widehat{\theta}^*) - \operatorname{Bias}^*(\widehat{\tau}^*)$ . Note that this bias corrected estimator requires the consistent estimators of  $\rho$  and  $\kappa$ . To that end, we start from a consistent estimator for  $\widetilde{\epsilon} = \widetilde{\epsilon}(\beta_0, \zeta_0) = \mathbb{Q}_{\mathbb{C}}(\tau_0) \mathbf{e}^{\tau_0 \mathbf{M}}(\mathbf{e}^{\alpha_0 \mathbf{W}}y - X\beta_0) = \mathbb{Q}_{\mathbb{C}}(\tau_0)\epsilon$ , which is given by  $\widehat{\epsilon} = \widehat{\epsilon}(\widehat{\beta}^*, \widehat{\zeta}^*) = \mathbb{Q}_{\mathbb{C}}(\widehat{\tau}^*) \mathbf{e}^{\widehat{\tau}^* \mathbf{M}}(\mathbf{e}^{\widehat{\alpha}^* \mathbf{W}}y - X\widehat{\beta}^*)$ . Let  $q_{jh}$  be the (j, h)th element of  $\mathbb{Q}_{\mathbb{C}}(\tau_0)$ . Let  $\epsilon_j$  and  $\widetilde{\epsilon}_j$  be the *j*th element of  $\epsilon$  and  $\widetilde{\epsilon}$ , respectively. Then,  $\widetilde{\epsilon}_j = q_{j1}\epsilon_1 + \ldots + q_{jN}\epsilon_N$ . Utilizing the fact that  $\epsilon_{it}$ 's are *i.i.d*, we have  $\mathrm{E}(\widetilde{\epsilon}_j^3) = \sum_{h=1}^N q_{jh}^3 \mathrm{E}(\epsilon_h^3) = \sigma_{\epsilon 0}^3 \rho \sum_{h=1}^N q_{jh}^3$ . Summing  $\mathrm{E}(\widetilde{\epsilon}_j^3)$  over *j* and solving for  $\rho$ , we obtain

$$\rho = \frac{\sum_{j=1}^{N} E(\tilde{\epsilon}_j^3)}{\sigma_{\epsilon 0}^3 \sum_{j=1}^{N} \sum_{h=1}^{N} q_{jh}^3}.$$
(3.20)

This result suggests the following sample counterpart as a consistent estimator for  $\rho$ :

$$\widehat{\rho} = \frac{\sum_{j=1}^{N} \widehat{\epsilon}_j^3}{\widehat{\sigma}_{\epsilon}^{*3} \sum_{j=1}^{N} \sum_{h=1}^{N} \widehat{q}_{jh}^3},\tag{3.21}$$

where  $\hat{\epsilon}_j$  is the *j*th element of  $\hat{\epsilon}(\hat{\beta}^*, \hat{\zeta}^*)$  and  $\hat{q}_{jh}$  is the (j, h)th element of  $\mathbb{Q}_{\mathbb{C}}(\hat{\tau}^*)$ . Similarly,

$$E(\tilde{\epsilon}_{j}^{4}) = \sum_{h=1}^{N} q_{jh}^{4} E(\epsilon_{h}^{4}) + 3\sigma_{\epsilon 0}^{4} \sum_{h=1}^{N} \sum_{l=1}^{N} q_{jh}^{2} q_{jl}^{2} - 3\sigma_{\epsilon 0}^{4} \sum_{h=1}^{N} q_{jh}^{4}$$
$$= \kappa \sum_{h=1}^{N} q_{jh}^{4} \sigma_{\epsilon 0}^{4} + 3\sigma_{\epsilon 0}^{4} \sum_{h=1}^{N} \sum_{l=1}^{N} q_{jh}^{2} q_{jl}^{2}.$$
(3.22)

Summing  $E(\tilde{\epsilon}_j^4)$  over j and solving for  $\kappa$ , we have

$$\kappa = \frac{\sum_{j=1}^{N} \mathrm{E}(\tilde{\epsilon}_{j}^{4}) - 3\sigma_{\epsilon 0}^{4} \sum_{j=1}^{N} \sum_{h=1}^{N} \sum_{l=1}^{N} q_{jh}^{2} q_{jl}^{2}}{\sigma_{\epsilon 0}^{4} \sum_{j=1}^{N} \sum_{h=1}^{N} q_{jh}^{4}}.$$
(3.23)

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

$$\widehat{\kappa} = \frac{\sum_{j=1}^{N} \widehat{\epsilon}_{j}^{4} - 3\widehat{\sigma}_{\epsilon}^{*4} \sum_{j=1}^{N} \sum_{h=1}^{N} \sum_{l=1}^{N} \widehat{q}_{jh}^{2} \widehat{q}_{jl}^{2}}{\widehat{\sigma}_{\epsilon}^{*4} \sum_{j=1}^{N} \sum_{h=1}^{N} \widehat{q}_{jh}^{4}}.$$
(3.24)

The next theorem shows that  $\hat{\rho}$  and  $\hat{\kappa}$  are consistent estimators.

**Theorem 3.4.** Under Assumptions 1–7, as  $N \to \infty$ , we have  $\hat{\rho} = \rho + o_p(1)$  and  $\hat{\kappa} = \kappa + o_p(1)$ .

Proof. See Section C in the web appendix.

### 4 Estimation under an unknown form of heteroskedasticity

In this section, we allow for both cross-sectional and time-series heteroskedasticity, and present a robust estimation and inference method for our matrix exponential unbalanced panel data model. Heteroskedasticity in matrix exponential models has been discussed in Debarsy et al. (2015) for cross-sectional data, and in Zhang et al. (2019) for balanced panel data. However in both papers the discussion on the (quasi) likelihood based estimation scheme has been limited to the case where the spatial weights matrix W involving the dependent variable and M involving the error terms are commutative, i.e., WM = MW. These papers show that the QMLE is consistent, if the commutativity property holds. However, this desirable feature of the matrix exponential specification extends neither to the matrix exponential unbalanced panel data models nor to the matrix exponential balanced panel data models with two-way error components. Therefore, a robust method for estimation and inference is highly desired for the matrix exponential unbalanced panel data models with an unknown form of heteroskedasticity. The following assumption specifies the form of heteroskedasticity in our setting.

**Assumption 8.** The  $\{\epsilon_j\}$  for j = 1, ..., N are independent but not identically distributed (i.n.i.d), i.e.,  $\{\epsilon_j\} \sim i.n.i.d(0, \sigma_j^2)$  and  $E |\epsilon_j|^{4+\varrho} < \infty$  for some  $\varrho > 0$ . Let  $\Gamma = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_N^2)$ , i.e.,  $\operatorname{Var}(\epsilon) = \Gamma$ . As in the case of homoskedasticity, we will compute the expectations of the  $\alpha$  and  $\tau$  elements of the concentrated quasi score function in (3.5) under Assumption 8, and then adjust these elements using their corresponding expectations. First, we consider the score function with respect to  $\alpha$ . Recall that  $\mathbf{G}(\zeta_0)y = \mathbf{e}^{\boldsymbol{\tau}_0\mathbf{M}}\phi + \epsilon$ ,  $\mathbf{S}(\tau_0) = \mathbf{e}^{\boldsymbol{\tau}_0\mathbf{M}}\mathbf{W}\mathbf{e}^{-\boldsymbol{\tau}_0\mathbf{M}}$ and  $\tilde{\epsilon} = \mathbb{Q}_{\mathbb{C}}(\tau_0)\epsilon$ . Then, using the fact that  $\operatorname{tr}(DA) = \operatorname{tr}(D\operatorname{diag}(A))$  for a diagonal matrix D and any conformable matrix A, we have

$$E\left(y'\mathbf{e}^{\boldsymbol{\alpha}_{0}\mathbf{W}'}\mathbf{W}'\mathbf{e}^{\boldsymbol{\tau}_{0}\mathbf{M}'}\tilde{\boldsymbol{\epsilon}}\right) = E\left(y'\mathbf{G}'(\zeta_{0})\mathbf{S}'(\tau_{0})\mathbb{Q}_{\mathbb{C}}(\tau_{0})\boldsymbol{\epsilon}\right) = \operatorname{tr}\left(\Gamma\operatorname{diag}\left(\mathbf{S}'(\tau_{0})\mathbb{Q}_{\mathbb{C}}(\tau_{0})\right)\right) = \operatorname{tr}\left(\Gamma\operatorname{diag}\left(\mathbf{S}'(\tau_{0})\mathbb{Q}_{\mathbb{C}}(\tau_{0})\right)\operatorname{diag}\left(\mathbb{Q}_{\mathbb{C}}(\tau_{0})\right)^{-1}\mathbb{Q}_{\mathbb{C}}(\tau_{0})\right), \quad (4.1)$$

where the last equality follows from the fact that  $\Gamma \operatorname{diag} \left( \mathbf{S}'(\tau_0) \mathbb{Q}_{\mathbb{C}}(\tau_0) \right) \operatorname{diag} \left( \mathbb{Q}_{\mathbb{C}}(\tau_0) \right)^{-1}$  is a diagonal matrix. Let  $\mathbf{\bar{S}}'(\tau_0) = \operatorname{diag} \left( \mathbf{S}'(\tau_0) \mathbb{Q}_{\mathbb{C}}(\tau_0) \right) \operatorname{diag} \left( \mathbb{Q}_{\mathbb{C}}(\tau_0) \right)^{-1}$ . Then, using  $\mathbf{G}(\zeta_0) y = \mathbf{e}^{\boldsymbol{\tau}_0 \mathbf{M}} \phi + \epsilon$ , the result in (4.1) suggests that

$$E\left(y'\mathbf{e}^{\boldsymbol{\alpha}_{0}\mathbf{W}'}\mathbf{W}'\mathbf{e}^{\boldsymbol{\tau}_{0}\mathbf{M}'}\tilde{\boldsymbol{\epsilon}}\right) = \operatorname{tr}\left(\Gamma\bar{\mathbf{S}}'(\tau_{0})\mathbb{Q}_{\mathbb{C}}(\tau_{0})\right) = E\left(y'\mathbf{G}'(\zeta_{0})\bar{\mathbf{S}}'(\tau_{0})\mathbb{Q}_{\mathbb{C}}(\tau_{0})\boldsymbol{\epsilon}\right),$$
  

$$\implies E\left(y'\mathbf{G}'(\zeta_{0})\mathbf{S}'(\tau_{0})\mathbb{Q}_{\mathbb{C}}(\tau_{0})\boldsymbol{\epsilon}\right) - E\left(y'\mathbf{G}'(\zeta_{0})\bar{\mathbf{S}}'(\tau_{0})\mathbb{Q}_{\mathbb{C}}(\tau_{0})\boldsymbol{\epsilon}\right) = 0$$
  

$$\implies E\left(y'\mathbf{G}'(\zeta_{0})\left(\mathbf{S}'(\tau_{0}) - \bar{\mathbf{S}}'(\tau_{0})\right)\mathbb{Q}_{\mathbb{C}}(\tau_{0})\boldsymbol{\epsilon}\right) = 0.$$
(4.2)

Thus, our suggested robust adjusted quasi score function for  $\alpha$  is the sample counterpart of the result in (4.2):

$$-y'\mathbf{G}'(\zeta)\left(\mathbf{S}'(\tau) - \bar{\mathbf{S}}'(\tau)\right)\tilde{\epsilon}(\beta,\zeta).$$
(4.3)

By construction, the quadratic form in (4.3) has an expectation of zero at the true parameter value  $\theta_0$  under Assumption 8. Using similar steps for the  $\tau$  element of the concentrated quasi score function in (3.5), we obtain

$$E\left(\tilde{\epsilon}'\mathbf{M}\tilde{\epsilon}\right) = E\left(\epsilon'\mathbb{Q}_{\mathbb{C}}(\tau_{0})\mathbf{M}\mathbb{Q}_{\mathbb{C}}(\tau_{0})\epsilon\right) = \operatorname{tr}\left(\Gamma\mathcal{Q}_{4}(\tau_{0})\right) = \operatorname{tr}\left(\Gamma\operatorname{diag}(\mathcal{Q}_{4}(\tau_{0}))\right) \\
 = \operatorname{tr}\left(\Gamma\operatorname{diag}\left(\mathcal{Q}_{4}(\tau_{0})\right)\operatorname{diag}\left(\mathbb{Q}_{\mathbb{C}}(\tau_{0})\right)^{-1}\mathbb{Q}_{\mathbb{C}}(\tau_{0})\right) = E\left(\epsilon'\overline{\mathcal{Q}}_{4}(\tau_{0})\mathbb{Q}_{\mathbb{C}}(\tau_{0})\epsilon\right),$$
(4.4)

where  $\overline{\mathcal{Q}}_4(\tau_0) = \operatorname{diag}(\mathcal{Q}_4(\tau_0))\operatorname{diag}(\mathbb{Q}_{\mathbb{C}}(\tau_0))^{-1}$ . A quadratic form that has a zero mean can be constructed from (4.4) as

$$E\left(\epsilon' \mathbb{Q}_{\mathbb{C}}(\tau_0) \mathbf{M} \mathbb{Q}_{\mathbb{C}}(\tau_0) \epsilon\right) - E\left(\epsilon' \overline{\mathcal{Q}}_4(\tau_0) \mathbb{Q}_{\mathbb{C}}(\tau_0) \epsilon\right) = 0 \implies E\left(\epsilon' \left(\mathcal{Q}_4(\tau_0) - \overline{\mathcal{Q}}_4(\tau_0)\right) \mathbb{Q}_{\mathbb{C}}(\tau_0) \epsilon\right) = 0.$$

$$(4.5)$$

Also, note that

$$\mathbf{E}\left(\boldsymbol{\epsilon}'\left(\mathcal{Q}_{4}(\tau_{0})-\overline{\mathcal{Q}}_{4}(\tau_{0})\right)\mathbb{Q}_{\mathbb{C}}(\tau_{0})\boldsymbol{\epsilon}\right)=0\implies \mathbf{E}\left(\left(\mathbf{e}^{\boldsymbol{\tau}_{0}\mathbf{M}}C\delta_{0}+\boldsymbol{\epsilon}\right)'\left(\mathcal{Q}_{4}(\tau_{0})-\overline{\mathcal{Q}}_{4}(\tau_{0})\right)\mathbb{Q}_{\mathbb{C}}(\tau_{0})\boldsymbol{\epsilon}\right)=0$$

Since  $\mathbf{e}^{\boldsymbol{\tau}_0 \mathbf{M}} \left( \mathbf{e}^{\boldsymbol{\alpha}_0 \mathbf{W}} y - X \beta \right) = \mathbf{e}^{\boldsymbol{\tau}_0 \mathbf{M}} C \delta_0 + \epsilon$ , the sample counterpart of (4.5) is given by

$$-(\mathbf{e}^{\boldsymbol{\alpha}\mathbf{W}}y - X\beta)'\mathbf{e}^{\boldsymbol{\tau}\mathbf{M}'}\left(\mathcal{Q}_4(\tau) - \overline{\mathcal{Q}}_4(\tau)\right)\tilde{\epsilon}(\beta,\zeta).$$

$$(4.6)$$

Then, using (4.3) and (4.6) together with the  $\beta$  element of the adjusted quasi score functions in (3.5), which is robust against the unknown heteroskedasticity, we suggest the following robust adjusted quasi score functions under Assumption 8:

$$S^{\dagger}(\beta,\zeta) = \begin{cases} \beta : & \mathbb{X}'(\tau)\tilde{\epsilon}(\beta,\zeta), \\ \alpha : & -y'\mathbf{G}'(\zeta)\left(\mathbf{S}'(\tau) - \mathbf{\bar{S}}'(\tau)\right)\tilde{\epsilon}(\beta,\zeta), \\ \tau : & -\left(\mathbf{e}^{\boldsymbol{\alpha}\mathbf{W}}y - X\beta\right)'\mathbf{e}^{\boldsymbol{\tau}\mathbf{M}'}\left(\mathcal{Q}_{4}(\tau) - \mathbf{\bar{Q}}_{4}(\tau)\right)\tilde{\epsilon}(\beta,\zeta). \end{cases}$$
(4.7)

Note that from (4.1), we have  $\frac{1}{N} \mathbb{E} \left( y' \mathbf{e}^{\alpha_0 \mathbf{W}'} \mathbf{W}' \mathbf{e}^{\tau_0 \mathbf{M}'} \tilde{\epsilon} \right) = \frac{1}{N} \operatorname{tr} \left( \Gamma \mathbf{S}'(\tau_0) \mathbb{Q}_{\mathbb{C}}(\tau_0) \right) = \frac{1}{N} \operatorname{tr} \left( \Gamma \mathbf{S}'(\tau_0) \right) - \frac{1}{N} \operatorname{tr} \left( \Gamma \mathbf{S}'(\tau_0) \mathbb{P}_{\mathbb{C}}(\tau_0) \right) = O(1/h_n) + O(1/\max\{h_n, T\}) = O(1/h_n) \text{ by Lemmas A.1 and A.2 of}$ the web appendix. Note that if  $\mathbf{M}$  and  $\mathbf{W}$  are commutative, then  $\operatorname{tr} \left( \Gamma \mathbf{S}'(\tau_0) \right) = 0$ , suggesting that  $\frac{1}{N} \mathbb{E} \left( y' \mathbf{e}^{\alpha_0 \mathbf{W}'} \mathbf{W}' \mathbf{e}^{\tau_0 \mathbf{M}'} \tilde{\epsilon} \right) = O(1/\max\{h_n, T\})$ . Also, from (4.4), we have  $\frac{1}{N} \mathbb{E} \left( \tilde{\epsilon}' \mathbf{M} \tilde{\epsilon} \right) = \frac{1}{N} \operatorname{tr} \left( \Gamma \mathcal{Q}_4(\tau_0) \right) = \frac{1}{N} \operatorname{tr} \left( \Gamma \mathbb{Q}_{\mathbb{C}}(\tau_0) \mathbf{M} \mathbb{Q}_{\mathbb{C}}(\tau_0) \right) = \frac{1}{N} \operatorname{tr} \left( \Gamma \mathbb{P}_{\mathbb{C}}(\tau_0) \mathbf{M} \mathbb{P}_{\mathbb{C}}(\tau_0) \right) - \frac{1}{N} \operatorname{tr} \left( \Gamma \mathbb{M} \mathbb{P}_{\mathbb{C}}(\tau_0) \right) - \frac{1}{N} \operatorname{tr} \left( \Gamma \mathbb{P}_{\mathbb{C}}(\tau_0) \mathbf{M} \right) = O(1/\max\{h_n, T\})$  by Lemmas A.1 and A.2 of the web appendix. Thus, under heteroskedasticity, the necessary condition  $\operatorname{plim}_{N \to \infty} \frac{1}{N} S^c(\theta_0) = 0$  for the consistency of the QMLE is satisfied if  $h_n$  is divergent. However, if  $h_n$  is bounded, then we should use  $S^{\dagger}(\beta, \zeta)$  for ensuring consistent estimation. To derive the robust M-estimator from (4.7), we can similarly first solve for  $\beta$  given  $\zeta$ , which is the same as (3.8):

$$\widehat{\boldsymbol{\beta}}^{\dagger}(\boldsymbol{\zeta}) = \widehat{\boldsymbol{\beta}}^{*}(\boldsymbol{\zeta}) = \left(\boldsymbol{X}' \mathbf{e}^{\boldsymbol{\tau} \mathbf{M}'} \mathbb{Q}_{\mathbb{C}}(\boldsymbol{\tau}) \mathbf{e}^{\boldsymbol{\tau} \mathbf{M}} \boldsymbol{X}\right)^{-1} \boldsymbol{X}' \mathbf{e}^{\boldsymbol{\tau} \mathbf{M}'} \mathbb{Q}_{\mathbb{C}}(\boldsymbol{\tau}) \mathbf{e}^{\boldsymbol{\tau} \mathbf{M}} \mathbf{e}^{\boldsymbol{\alpha} \mathbf{W}} \boldsymbol{y}.$$
(4.8)

Then, substituting  $\hat{\beta}^{\dagger}(\zeta)$  into the  $\alpha$  and  $\tau$  elements of (4.7), we obtain the concentrated robust adjusted quasi score functions as:

$$S^{\dagger c}(\zeta) = \begin{cases} \alpha : & -y' \mathbf{G}'(\zeta) \left( \mathbf{S}'(\tau) - \overline{\mathbf{S}}'(\tau) \right) \widehat{\epsilon}(\zeta), \\ \tau : & - \left( \mathbf{e}^{\alpha \mathbf{W}} y - X \widehat{\beta}^{\dagger}(\zeta) \right)' \mathbf{e}^{\tau \mathbf{M}'} \left( \mathcal{Q}_4(\tau) - \overline{\mathcal{Q}}_4(\tau) \right) \widehat{\epsilon}(\zeta), \end{cases}$$
(4.9)

where  $\hat{\epsilon}(\zeta) = \tilde{\epsilon}(\hat{\beta}^{\dagger}(\zeta), \zeta)$ . Then, our robust M-estimator (RME)  $\hat{\zeta}^{\dagger}$  of  $\zeta_0$  is defined by  $\hat{\zeta}^{\dagger} = \arg \operatorname{solve}\{S^{\dagger c}(\zeta) = 0\}$ . Substituting  $\hat{\zeta}^{\dagger}$  into (4.8), we will get the RME  $\hat{\beta}^{\dagger} = \hat{\beta}^{\dagger}(\hat{\zeta}^{\dagger})$  of  $\beta_0$ .

Similar to the homoskedasticity case, we will prove the consistency of  $\hat{\zeta}^{\dagger}$ , which leads to the consistency of  $\hat{\beta}^{\dagger}(\hat{\zeta}^{\dagger})$ . Let  $\bar{S}^{\dagger}(\beta,\zeta) = E(S^{\dagger}(\beta,\zeta))$  be the population counterpart of the robust adjusted quasi score functions. Given  $\zeta$ , we can derive the estimator  $\bar{\beta}^{\dagger}(\zeta) = \bar{\beta}^{*}(\zeta)$  in (3.11), which can be substituted into the  $\alpha$  and  $\tau$  elements of  $\bar{S}^{\dagger}(\beta,\zeta)$  to obtain:

$$\bar{S}^{\dagger c}(\zeta) = \begin{cases} \alpha : & -\mathrm{E}\left(y'\mathbf{G}'(\zeta)\left(\mathbf{S}'(\tau) - \bar{\mathbf{S}}'(\tau)\right)\bar{\epsilon}(\zeta)\right), \\ \tau : & -\mathrm{E}\left(\left(\mathbf{e}^{\alpha \mathbf{W}}y - X\bar{\beta}^{\dagger}(\zeta)\right)'\mathbf{e}^{\tau \mathbf{M}'}\left(\mathcal{Q}_{4}(\tau) - \bar{\mathcal{Q}}_{4}(\tau)\right)\bar{\epsilon}(\zeta)\right), \end{cases}$$
(4.10)

where  $\bar{\epsilon}(\zeta) = \tilde{\epsilon}(\bar{\beta}^{\dagger}(\zeta), \zeta)$ . Then, similar to the homoskedastic case,  $\hat{\zeta}^{\dagger}$  is a consistent estimator if  $\sup_{\zeta \in \Delta} \frac{1}{N_1} \left\| S^{\dagger c}(\zeta) - \bar{S}^{\dagger c}(\zeta) \right\| \xrightarrow{p} 0$  and the following identification assumption holds.

**Assumption 9.**  $\inf_{\zeta: d(\zeta,\zeta_0) \ge \vartheta} \|\overline{S}^{\dagger c}(\zeta)\| > 0$  for every  $\vartheta > 0$ , where  $d(\zeta,\zeta_0)$  is a measure of distance between  $\zeta$  and  $\zeta_0$ .

Let  $\omega = (\beta', \zeta')'$  and  $\widehat{\omega}^{\dagger} = (\widehat{\beta}^{\dagger'}, \widehat{\zeta}^{\dagger'})'$ , then we have the following theorem.

**Theorem 4.1.** Under Assumptions 2–5 and 7–9, as  $N \to \infty$ , we have  $\widehat{\omega}^{\dagger} \xrightarrow{p} \omega_0$ .

*Proof.* See Section C in the web appendix.

Substituting  $\tilde{\epsilon}(\beta_0, \zeta_0) = \mathbb{Q}_{\mathbb{C}}(\tau_0)\epsilon$  and  $y = \mathbf{e}^{-\boldsymbol{\alpha}_0 \mathbf{W}} \left(\phi + \mathbf{e}^{-\boldsymbol{\tau}_0 \mathbf{M}}\epsilon\right)$  into  $S^{\dagger}(\omega_0)$ , we have

$$S^{\dagger}(\omega_{0}) = \begin{cases} \beta : \quad \mathbb{X}'(\tau_{0})\epsilon, \\ \alpha : \quad -\phi' \mathbf{e}^{\boldsymbol{\tau}_{0}\mathbf{M}'} \left(\mathbf{S}'(\tau_{0}) - \mathbf{\bar{S}}'(\tau_{0})\right) \mathbb{Q}_{\mathbb{C}}(\tau_{0})\epsilon - \epsilon' \left(\mathbf{S}'(\tau_{0}) - \mathbf{\bar{S}}'(\tau_{0})\right) \mathbb{Q}_{\mathbb{C}}(\tau_{0})\epsilon, \\ \tau : \quad -\delta'_{0}\mathbb{C}'(\tau_{0}) \left(\mathcal{Q}_{4}(\tau_{0}) - \mathbf{\bar{Q}}_{4}(\tau_{0})\right) \mathbb{Q}_{\mathbb{C}}(\tau_{0})\epsilon - \epsilon' \left(\mathcal{Q}_{4}(\tau_{0}) - \mathbf{\bar{Q}}_{4}(\tau_{0})\right) \mathbb{Q}_{\mathbb{C}}(\tau_{0})\epsilon. \end{cases}$$

$$(4.11)$$

Similar to the homoskedastic case, we can show the asymptotic normality of  $\frac{1}{\sqrt{N_1}}S^{\dagger}(\omega_0)$  using the CLT for linear-quadratic forms in Kelejian and Prucha (2001). Together with the appropriate asymptotic property of  $\frac{1}{N_1}\frac{\partial S^{\dagger}(\overline{\omega})}{\partial \omega'}$  for some  $\overline{\omega}$  between  $\widehat{\omega}^{\dagger}$  and  $\omega_0$  elementwise, we have the following theorem.

**Theorem 4.2.** Under Assumptions 2–5 and 7–9, as  $N \to \infty$ ,

$$\sqrt{N_1} \left( \widehat{\omega}^{\dagger} - \omega_0 \right) \stackrel{d}{\longrightarrow} N \left( 0, \lim_{n \to \infty} \Psi^{\dagger - 1}(\omega_0) \Omega^{\dagger}(\omega_0) \Psi^{\dagger - 1'}(\omega_0) \right), \tag{4.12}$$

where  $\Psi^{\dagger}(\omega_0) = -\frac{1}{N_1} \operatorname{E}\left(\frac{\partial S^{\dagger}(\omega_0)}{\partial \omega'}\right)$  and  $\Omega^{\dagger}(\omega_0) = \frac{1}{N_1} \operatorname{Var}\left(S^{\dagger}(\omega_0)\right)$  are assumed to exist and  $\Psi^{\dagger}(\omega_0)$  is assumed to be positive definite for sufficiently large N.

*Proof.* See Section C in the web appendix.

We need consistent estimators of  $\Psi^{\dagger}(\omega_0)$  and  $\Omega^{\dagger}(\omega_0)$  to conduct valid inference. Given the expressions of  $\frac{\partial S^{\dagger}(\omega)}{\partial \omega'}$  in the proof of Theorem 4.2,  $\Psi^{\dagger}(\omega_0)$  can be estimated by its sample counterpart  $\Psi^{\dagger}(\hat{\omega}^{\dagger}) = -\frac{1}{N_1} \frac{\partial S^{\dagger}(\omega)}{\partial \omega'}|_{\omega=\hat{\omega}^{\dagger}}$ . Let  $H^{\dagger}_a(\omega) = -\frac{1}{N_1} \frac{\partial S^{\dagger}(\omega)}{\partial a'}$  for  $a \in \{\beta, \alpha, \tau\}$ . Then, we have

$$\begin{split} N_{1}H_{\beta\beta}^{\dagger}(\omega) &= \mathbb{X}'(\tau)\mathbb{X}(\tau), \quad N_{1}H_{\beta\alpha}^{\dagger}(\omega) = -\mathbb{X}'(\tau)y(\zeta), \quad N_{1}H_{\beta\tau}^{\dagger}(\omega) = -\mathbb{X}'(\tau)\mathbf{M}^{s}\tilde{\epsilon}(\beta,\zeta), \\ N_{1}H_{\alpha\beta}^{\dagger}(\omega) &= -y'\mathbf{G}'(\zeta)\left(\mathbf{S}'(\tau) - \mathbf{\bar{S}}'(\tau)\right)\mathbb{X}(\tau), \\ N_{1}H_{\alpha\alpha}^{\dagger}(\omega) &= y'\mathbf{G}'(\zeta)\left(\mathbf{S}'(\tau) - \mathbf{\bar{S}}'(\tau)\right)\tilde{\epsilon}(\beta,\zeta) + y'\mathbf{G}'(\zeta)\left(\mathbf{S}'(\tau) - \mathbf{\bar{S}}'(\tau)\right)y(\zeta), \\ N_{1}H_{\alpha\tau}^{\dagger}(\omega) &= y'\mathbf{G}'(\zeta)\left(\mathbf{M}'\left((\mathbf{S}'(\tau) - \mathbf{\bar{S}}'(\tau)\right) + \mathbf{\dot{S}}'(\tau) - \mathbf{\dot{\bar{S}}}'(\tau) + \left((\mathbf{S}'(\tau) - \mathbf{\bar{S}}'(\tau)\right)\mathbb{D}(\tau)\right)\tilde{\epsilon}(\beta,\zeta), \\ N_{1}H_{\tau\beta}^{\dagger}(\omega) &= -\tilde{\epsilon}'(\beta,\zeta)\left(\mathcal{Q}_{4}'(\tau) - \mathbf{\bar{Q}}_{4}'(\tau)\right)\mathbf{e}^{\tau\mathbf{M}}X - \left(\mathbf{e}^{\alpha\mathbf{W}}y - X\beta\right)'\mathbf{e}^{\tau\mathbf{M}'}\left(\mathcal{Q}_{4}(\tau) - \mathbf{\bar{Q}}_{4}(\tau)\right)\mathbb{X}(\tau), \\ N_{1}H_{\tau\alpha}^{\dagger}(\omega) &= y'\mathbf{W}'\mathbf{G}'(\zeta)\left(\mathcal{Q}_{4}(\tau) - \mathbf{\bar{Q}}_{4}(\tau)\right)\tilde{\epsilon}(\beta,\zeta) + \left(\mathbf{e}^{\alpha\mathbf{W}}y - X\beta\right)'\mathbf{e}^{\tau\mathbf{M}'}\left(\mathcal{Q}_{4}(\tau) - \mathbf{\bar{Q}}_{4}(\tau)\right)y(\zeta), \\ N_{1}H_{\tau\tau}^{\dagger}(\omega) &= \left(\mathbf{e}^{\alpha\mathbf{W}}y - X\beta\right)'\mathbf{e}^{\tau\mathbf{M}'}\left(\mathbf{M}'\left(\mathcal{Q}_{4}(\tau) - \mathbf{\bar{Q}}_{4}(\tau)\right) + \dot{\mathcal{Q}}_{4}(\tau) - \mathbf{\bar{Q}}_{4}(\tau) + \left(\mathcal{Q}_{4}(\tau) - \mathbf{\bar{Q}}_{4}(\tau)\right)\mathbb{D}(\tau)\right)\tilde{\epsilon}, \end{split}$$

where  $\dot{\mathbf{S}}'(\tau) = \operatorname{diag}\left(\dot{\mathbf{S}}'(\tau)\mathbb{Q}_{\mathbb{C}}(\tau) - \mathbf{S}'(\tau)\dot{\mathbb{Q}}_{\mathbb{C}}(\tau)\right)\operatorname{diag}\left(\mathbb{Q}_{\mathbb{C}}(\tau)\right)^{-1} + \mathbf{\bar{S}}(\tau)\operatorname{diag}\left(\dot{\mathbb{Q}}_{\mathbb{C}}(\tau)\right)\operatorname{diag}\left(\mathbb{Q}_{\mathbb{C}}(\tau)\right)^{-1},$  $\dot{\mathbf{S}}'(\tau) = \mathbf{S}'(\tau)\mathbf{M}' - \mathbf{M}'\mathbf{S}'(\tau), \ \dot{\overline{\mathcal{Q}}}_{4}(\tau) = \operatorname{diag}\left(\dot{\mathcal{Q}}_{4}(\tau)\right)\operatorname{diag}\left(\mathbb{Q}_{\mathbb{C}}(\tau)\right)^{-1} - \overline{\mathcal{Q}}_{4}(\tau)\operatorname{diag}\left(\dot{\mathbb{Q}}_{\mathbb{C}}(\tau)\right)\operatorname{diag}\left(\mathbb{Q}_{\mathbb{C}}(\tau)\right)^{-1},$  $\dot{\mathcal{Q}}_{4}(\tau) = \dot{\mathbb{Q}}_{\mathbb{C}}(\tau)\mathbf{M}\mathbb{Q}_{\mathbb{C}}(\tau) + \mathbb{Q}_{\mathbb{C}}(\tau)\mathbf{M}\dot{\mathbb{Q}}_{\mathbb{C}}(\tau) \text{ and } \dot{\mathbb{Q}}_{\mathbb{C}}(\tau) = -\left(\mathbb{Q}_{\mathbb{C}}(\tau)\mathbf{M}\mathbb{P}_{\mathbb{C}}(\tau) + \mathbb{P}_{\mathbb{C}}(\tau)\mathbf{M}'\mathbb{Q}_{\mathbb{C}}(\tau)\right).$  The proof for the consistency of  $\Psi^{\dagger}(\hat{\omega}^{\dagger})$  is provided in the proof of Theorem 4.2.

For  $\Omega^{\dagger}(\omega_0)$ , we first derive its closed-form expression using Lemma A.7 of the web appendix:

$$\begin{split} N_{1}\Omega_{\beta\omega}^{\dagger}(\omega_{0}) &= N_{1}\left(\Omega_{\beta\beta}^{\dagger}(\omega_{0}), \quad \Omega_{\beta\alpha}^{\dagger}(\omega_{0}), \quad \Omega_{\beta\tau}^{\dagger}(\omega_{0})\right) \\ &= \left(\mathbb{X}'(\tau_{0})\Gamma\mathbb{X}(\tau_{0}), \quad -\mathbb{X}'(\tau_{0})\Gamma\bar{\mathbf{S}}(\tau_{0})\mathbf{e}^{\boldsymbol{\tau}_{0}\mathbf{M}}\phi, \quad -\mathbb{X}'(\tau_{0})\Gamma\bar{\mathbf{Q}}(\tau_{0})\mathbb{C}(\tau_{0})\delta_{0}\right), \\ N_{1}\Omega_{\alpha\alpha}^{\dagger}(\omega_{0}) &= \phi'\mathbf{e}^{\boldsymbol{\tau}_{0}\mathbf{M}'}\bar{\mathbf{S}}'(\tau_{0})\Gamma\bar{\mathbf{S}}(\tau_{0})\mathbf{e}^{\boldsymbol{\tau}_{0}\mathbf{M}}\phi + \operatorname{tr}\left(\Gamma\bar{\mathbf{S}}(\tau_{0})\Gamma\bar{\mathbf{S}}^{s}(\tau_{0})\right), \\ N_{1}\Omega_{\alpha\tau}^{\dagger}(\omega_{0}) &= \phi'\mathbf{e}^{\boldsymbol{\tau}_{0}\mathbf{M}'}\bar{\mathbf{S}}'(\tau_{0})\Gamma\bar{\mathbf{Q}}(\tau_{0})\mathbb{C}(\tau_{0})\delta_{0} + \operatorname{tr}\left(\Gamma\bar{\mathbf{S}}(\tau_{0})\Gamma\bar{\mathbf{Q}}^{s}(\tau_{0})\right), \\ N_{1}\Omega_{\tau\tau}^{\dagger}(\omega_{0}) &= \delta_{0}'\mathbb{C}'(\tau_{0})\bar{\mathbf{Q}}'(\tau_{0})\Gamma\bar{\mathbf{Q}}(\tau_{0})\mathbb{C}(\tau_{0})\delta_{0} + \operatorname{tr}\left(\Gamma\bar{\mathbf{Q}}(\tau_{0})\Gamma\bar{\mathbf{Q}}^{s}(\tau_{0})\right), \end{split}$$

where  $\overline{\mathbf{S}}(\tau_0) = \mathbb{Q}_{\mathbb{C}}(\tau_0) \left( \mathbf{S}(\tau_0) - \overline{\mathbf{S}}(\tau_0) \right)$  and  $\overline{\overline{\mathcal{Q}}}(\tau_0) = \mathbb{Q}_{\mathbb{C}}(\tau_0) \left( \mathcal{Q}'_4(\tau_0) - \overline{\mathcal{Q}}'_4(\tau_0) \right)$ . When deriving the above terms we utilized the fact that diag  $\left( \overline{\mathbf{S}}(\tau_0) \right) = \text{diag} \left( \mathbb{Q}_{\mathbb{C}}(\tau_0) (\mathbf{S}(\tau_0) - \overline{\mathbf{S}}(\tau_0)) \right) = \text{diag} \left( \mathbb{Q}_{\mathbb{C}}(\tau_0) \mathbf{S}(\tau_0) \right) - \text{diag} \left( \mathbb{Q}_{\mathbb{C}}(\tau_0) \right)^{-1} \text{diag} \left( \mathbb{Q}_{\mathbb{C}}(\tau_0) \mathbf{S}(\tau_0) \right) = 0_{N \times N}$ , which implies that the diagonal elements of  $\overline{\mathbf{S}}(\tau_0)$  are zeroes. Similarly, the diagonal elements of  $\overline{\overline{\mathcal{Q}}}(\tau_0)$  are also zeroes. By Lemma A.7 of the web appendix, all terms involving the 3rd and 4th moments of the disturbances disappear, i.e.,  $\Omega^{\dagger}(\omega_0)$  is free from the 3rd and 4th moments of the disturbances. For convenience of exposition, let us write  $\Omega^{\dagger}(\omega_0)$  as  $\Omega^{\dagger}(\omega_0, \delta_0, \Gamma)$ . Let  $\widehat{\delta}^{\dagger}$  be the estimator of  $\delta_0$  derived from the RME using (3.3). Let  $\Omega^{\dagger}(\widehat{\omega}^{\dagger}, \widehat{\delta}^{\dagger}, \Gamma)$  be the estimator of  $\Omega^{\dagger}(\omega_0)$  given  $\Gamma$ . Then, we can derive a bias-corrected estimator for  $\Omega^{\dagger}(\omega_0, \delta_0, \Gamma)$  as follows.

Theorem 4.3. Under Assumptions 2-5 and 7-9, we have

$$\Omega^{\dagger}(\widehat{\omega}^{\dagger},\widehat{\delta}^{\dagger},\Gamma) = \Omega^{\dagger}(\omega_{0},\delta_{0},\Gamma) + \operatorname{Bias}_{\delta}^{\dagger}(\tau_{0},\Gamma) + o_{p}(1),$$

where  $\operatorname{Bias}^{\dagger}_{\delta}(\tau_0, \Gamma)$  is an  $(k+2) \times (k+2)$  matrix with zero entries everywhere except the  $\zeta$  part of

the matrix, which is given by  $\begin{pmatrix} \operatorname{Bias}_{\delta,\alpha\alpha}^{\dagger}(\tau_{0},\Gamma) & \operatorname{Bias}_{\delta,\alpha\tau}^{\dagger}(\tau_{0},\Gamma) \\ \operatorname{Bias}_{\delta,\tau\alpha}^{\dagger}(\tau_{0},\Gamma) & \operatorname{Bias}_{\delta,\tau\tau}^{\dagger}(\tau_{0},\Gamma) \end{pmatrix}, \text{ where }$ 

$$\begin{aligned} \operatorname{Bias}_{\delta,\alpha\alpha}^{\dagger}(\tau_{0},\Gamma) &= \frac{1}{N_{1}} \operatorname{tr} \left( \Gamma \mathbb{P}_{\mathbb{C}}(\tau_{0}) \overline{\overline{\mathbf{S}}}'(\tau_{0}) \Gamma \overline{\overline{\mathbf{S}}}(\tau_{0}) \mathbb{P}_{\mathbb{C}}(\tau_{0}) \right), \\ \operatorname{Bias}_{\delta,\alpha\tau}^{\dagger}(\tau_{0},\Gamma) &= \operatorname{Bias}_{\delta,\tau\alpha}^{\dagger}(\tau_{0},\Gamma) = \frac{1}{N_{1}} \operatorname{tr} \left( \Gamma \mathbb{P}_{\mathbb{C}}(\tau_{0}) \overline{\overline{\mathbf{Q}}}'(\tau_{0}) \Gamma \overline{\overline{\mathbf{S}}}(\tau_{0}) \mathbb{P}_{\mathbb{C}}(\tau_{0}) \right), \text{ and} \\ \operatorname{Bias}_{\delta,\alpha\alpha}^{\dagger}(\tau_{0},\Gamma) &= \frac{1}{N_{1}} \operatorname{tr} \left( \Gamma \mathbb{P}_{\mathbb{C}}(\tau_{0}) \overline{\overline{\mathbf{Q}}}'(\tau_{0}) \Gamma \overline{\overline{\mathbf{Q}}}(\tau_{0}) \mathbb{P}_{\mathbb{C}}(\tau_{0}) \right). \end{aligned}$$

*Proof.* See Section C in the web appendix.

Under our assumptions and Lemma A.2 of the web appendix, it can be shown that  $\operatorname{Bias}^{\dagger}_{\delta}(\tau_0, \Gamma) = O(1/\max\{T, h_n\})$ . Therefore, in settings with large T and bounded  $h_n$ , or with fixed T and divergent  $h_n$ , the bias term becomes negligible. However, in settings with fixed T and bounded  $h_n$ , the bias correction is necessary for valid inference. Now we need to provide consistent estimators for the trace terms involving  $\Gamma$ . Since  $\tilde{\epsilon} = \tilde{\epsilon}(\beta_0, \zeta_0) = \mathbb{Q}_{\mathbb{C}}(\tau_0)\epsilon$ , we have  $\mathrm{E}(\tilde{\epsilon} \odot \tilde{\epsilon}) = (\mathbb{Q}_{\mathbb{C}}(\tau_0) \odot \mathbb{Q}_{\mathbb{C}}(\tau_0)) (\sigma_1^2, \ldots, \sigma_N^2)'$ , where  $\odot$  denotes the Hadamard product. This implies an estimator for  $(\sigma_1^2, \ldots, \sigma_N^2)'$  as

$$\left(\widehat{\sigma}_{1}^{2},\ldots,\widehat{\sigma}_{N}^{2}\right)' = \left(\mathbb{Q}_{\mathbb{C}}(\widehat{\tau}^{\dagger}) \odot \mathbb{Q}_{\mathbb{C}}(\widehat{\tau}^{\dagger})\right)^{-} (\widehat{\epsilon} \odot \widehat{\epsilon}), \tag{4.13}$$

where  $A^-$  denotes the generalized inverse of A and  $\hat{\epsilon} = \hat{\epsilon}(\hat{\beta}^{\dagger}, \hat{\zeta}^{\dagger}) = \mathbb{Q}_{\mathbb{C}}(\hat{\tau}^{\dagger})\mathbf{e}^{\hat{\tau}^{\dagger}\mathbf{M}}(\mathbf{e}^{\hat{\alpha}^{\dagger}\mathbf{W}}y - X\hat{\beta}^{\dagger}).$ Note that elements of  $\Omega^{\dagger}(\omega_0)$  takes forms of either  $\operatorname{tr}(\Gamma C)$  or  $\operatorname{tr}(\Gamma A \Gamma B)$ . We now show the effect of replacing the unknown  $\Gamma$  by  $\hat{\Gamma} = \operatorname{diag}(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_N^2)$  in large samples.

**Theorem 4.4.** Assume  $\Pi(\tau) = (\mathbb{Q}_{\mathbb{C}}(\tau) \odot \mathbb{Q}_{\mathbb{C}}(\tau))^{-1}$  exists for  $\tau$  in a neighborhood of  $\tau_0$  and is bounded in row and column sum norms. Let A and B be two  $N \times N$  matrices that have zero diagonal elements with elements  $a_{ij}$  and  $b_{ij}$ , respectively, and be bounded in row and column sum norms. Let C be an  $N \times N$  matrix that has uniformly bounded diagonal elements. Then,

(i) 
$$\frac{1}{N} \operatorname{tr}(\widehat{\Gamma}C) - \frac{1}{N} \operatorname{tr}(\Gamma C) = o_p(1),$$
  
(ii)  $\frac{1}{N} \operatorname{tr}(\widehat{\Gamma}A\widehat{\Gamma}B) - \frac{1}{N} \operatorname{tr}(\Gamma A\Gamma B) - \frac{2}{N} \operatorname{tr}((A \odot B)\Pi(\tau_0)\Lambda(\Gamma)\Pi(\tau_0)) = o_p(1),$ 

where  $\Lambda(\Gamma) = (\mathbb{Q}_{\mathbb{C}}(\tau_0)\Gamma\mathbb{Q}_{\mathbb{C}}(\tau_0)) \odot (\mathbb{Q}_{\mathbb{C}}(\tau_0)\Gamma\mathbb{Q}_{\mathbb{C}}(\tau_0)).$ 

*Proof.* See Section C in the web appendix.

Our results in Theorems 4.3 and 4.4 suggest that the bias corrected estimator of  $\Omega^{\dagger}(\omega_0)$  is given by

$$\widehat{\Omega}^{\dagger} = \Omega^{\dagger}(\widehat{\omega}^{\dagger}, \widehat{\delta}^{\dagger}, \widehat{\Gamma}) - \operatorname{Bias}_{\delta}^{\dagger}(\widehat{\tau}^{\dagger}, \widehat{\Gamma}) - \operatorname{Bias}_{\Gamma}^{\dagger}(\widehat{\tau}^{\dagger}, \widehat{\Gamma}),$$
(4.14)

where  $\operatorname{Bias}_{\Gamma}^{\dagger}(\widehat{\tau}^{\dagger},\widehat{\Gamma})$  is an  $(k+2)\times(k+2)$  matrix with zero entries everywhere except the  $\zeta$  part, which is given by  $\operatorname{Bias}_{\Gamma,ab}^{\dagger}(\tau_0) = \frac{2}{N_1} \operatorname{tr}\left(\left(\Xi_a(\tau_0)\odot\Xi_b^s(\tau_0) - \mathbb{P}_{\mathbb{C}}(\tau_0)\Xi_a^\prime(\tau_0)\odot\Xi_b(\tau_0)\mathbb{P}_{\mathbb{C}}(\tau_0)\right)\Pi(\tau_0)\Lambda(\Gamma)\Pi(\tau_0)\right)$ , where  $\Xi_{\alpha}(\tau_0) = \overline{\mathbf{S}}(\tau_0), \ \Xi_{\tau}(\tau_0) = \overline{\overline{\mathcal{Q}}}(\tau_0)$  and  $a, b \in \{\alpha, \tau\}$ . In the next corollary, we establish the consistency of  $\widehat{\Omega}^{\dagger}$  and  $\Psi^{\dagger}(\widehat{\omega}^{\dagger})$ .

**Corollary 4.1.** Under Assumptions 2–5 and 7–9, as  $N \to \infty$ ,  $\Psi^{\dagger}(\widehat{\omega}^{\dagger}) - \Psi^{\dagger}(\omega_0) \xrightarrow{p} 0$  and  $\widehat{\Omega}^{\dagger} - \Omega^{\dagger}(\omega_0) \xrightarrow{p} 0$ , and therefore  $\Psi^{\dagger-1}(\widehat{\omega}^{\dagger})\widehat{\Omega}^{\dagger}\Psi^{\dagger-1'}(\widehat{\omega}^{\dagger}) - \Psi^{\dagger-1}(\omega_0)\Omega^{\dagger}(\omega_0)\Psi^{\dagger-1'}(\omega_0) \xrightarrow{p} 0$ .

*Proof.* See Section D in the web appendix.

### 5 A Monte Carlo study

In this section, we investigate the finite sample performance of the proposed ME and RME. To this end, we consider the following specification for n entities with elements randomly missing over Tperiods,

$$e^{\alpha_0 W_t} y_t = X_{1t} \beta_{10} + X_{2t} \beta_{20} + C_t \mu_0 + \lambda_{t0} l_{n_t} + u_t, \quad e^{\tau_0 M_t} u_t = \epsilon_t, \quad t = 1, 2, \dots, T.$$
(5.1)

We first generate n elements for  $X_{1t}$  and  $X_{2t}$  by drawing independently from the standard normal distribution and  $U(0, \sqrt{12})$ , respectively, where  $U(0, \sqrt{12})$  denotes the uniform distribution  $(0, \sqrt{12})$ . Some of these observations are dropped subsequently when we introduce the unbalancedness. For the entity fixed effects, we consider a Mundlak type specification,  $\mu_0 = \frac{1}{T} \sum_{t=1}^T X_{1t} + e_t$ , where the elements of  $e_t$  are drawn independently from the standard normal distribution. The elements of the  $\lambda_{t0}$  are drawn independently from the standard normal distribution.

For the spatial weights matrices, we consider the rook and the queen contiguity cases. To this end, we first generate a vector containing a random permutation of the integers from 1 to n without

repeating elements. Then, we reshape this vector into an  $k \times m$  rectangular lattice, where m = n/k. In the rook contiguity case, we set the ij-th element of the weights matrix to 1 if the j'th entity is adjacent (left/right/above or below) to the i'th entity on the lattice. In the queen contiguity case, we set the ij-th element of the weights matrix to 1 if the j'th entity is adjacent to, or shares a border with the i'th entity. Subsequently, the rows and columns corresponding to the missing entities are deleted from  $W_t$  and  $M_t$  when we introduce the unbalancedness, and then we row normalize the weights matrices.

We set  $\epsilon_{it} = \sigma_{it}\eta_{it}$ , and consider two cases for  $\eta_{it}$ . In the homoskedastic scenario, we set  $\sigma_{it} = 1$ for all *i* and *t*, and generate  $\eta_{it}$ 's independently from either (i) the standard normal distribution or (ii) the standardized Gamma(2, 1), where Gamma(*a*, *b*) denotes the gamma distribution with the shape parameter *a* and the scale parameter *b*. In the heteroskedastic scenario, we consider  $\sigma_{it}^2 = \exp(0.1 + 0.35X_{2,it})$ , where  $X_{2,it}$  is the (*i*, *t*)th element of  $X_2$ , and  $\eta_{it}$ 's are again drawn independently from either (i) the standard normal distribution or (ii) the standardized Gamma(2, 1). In the accompanying web appendix, we consider two additional specifications for the variance terms.

The unbalancedness of the panels is introduced by randomly dropping at most 10% of the n entities in each period, while making sure that each entity is observed at least twice over T periods. For n and T we consider the following values,  $n \in \{50, 100\}$  and  $T \in \{3, 7\}$ . The true parameter values are  $\beta_{10} = -1$ ,  $\beta_{20} = 2$ ,  $\alpha_0 = -2$ ,  $\tau_0 \in \{-1, 1\}$ . The number of resamples is 1000 in each experiment. The matrix-vector products approach considered in Yang et al. (2021) can be used for fast estimation of our model. We provide pseudo estimation algorithms describing the steps in the M-estimation in the accompanying web appendix.

The results from the experiments are presented in Tables 1 - 4. In these tables, we report the empirical mean, the empirical standard deviation (in brackets) and the average of the asymptotic standard error (in square brackets) for the QMLE, ME and RME. The asymptotic standard errors are calculated using the consistent estimators of the variance-covariance matrices of the ME and RME provided in Theorems 3.2 and 4.2, respectively. Below, we summarize our findings from the tables.

1. As expected, the QMLE suffers from the incidental parameter problem. When T = 3, the QMLE of  $\sigma^2$  is downward biased significantly, close to -40%. As T increases to 7, the bias

gets smaller, but there still remains a significant amount of bias, about -20%.

- 2. In general, the ME and the RME show excellent finite sample performance in terms of finite sample bias. This result is not sensitive to the presence of heteroskedasticity and/or the type of distribution chosen for the disturbances. For example, in Table 1 when T = 3 and τ<sub>0</sub> = -1, we observe that the bias of the ME and the RME are very close to each other for all parameters. The largest of the bias values occur for τ in the case where n = 50, W=Queen and M=Rook. In this case, the bias of the ME is about -2.6% and the bias of the RME is about -2.5%. When T increases to 7, we observe that the bias of the ME and the ME and the Bias of the RME is about -2.5%. When T increases to 7, we observe that the bias of the ME and RME impose almost no bias.
- 3. Similarly, in the small T case, when  $\tau_0 = 1$ , we observe that the bias of the ME and the RME are very close to each other for all parameters. The largest of the bias values occur for  $\tau$  in the case where W=Rook and M=Queen. For example, in Table 3 for n = 50, the bias of the ME is about 7.2% and the bias of the RME is about 7.8%. It gets significantly smaller when T increases to 7, and we observe that the bias of the ME and the bias of the RME decrease to about 2%.
- 4. In terms of finite sample inference, we observe that both the ME and the RME show excellent performance. The reported average values for the asymptotic standard errors are very close to the corresponding empirical standard deviations for all parameters. We also observe that as T increases both the ME and the RME become more precise as both report smaller empirical standard deviations and average asymptotic standard errors. These findings are not sensitive to the choice of weights matrices, and/or the distributional specification for the disturbance term.
- 5. In the heteroskedastic scenario, we expect that the RME should outperform the ME in terms of finite sample bias and inference. However, the results from Tables 2 and 4 show that our heteroskedasticity specification using the skedastic function does not seem to deteriorate the finite sample properties of the ME significantly.

	W=Rook, M=Queen				W=Queen, M=Rook	
	QMLE	ME	RME	QMLE	ME	RME
			n = 50	, T = 3		
$\beta_{10} = -1$	-1.0008(.095)	-0.9996(.095)[.093]	-0.9995(.095)[.090]	-0.9982(.083)	-0.9997(.083)[.081]	-0.9999(.083)[.079]
$\beta_{20} = 2$	1.9962(.090)	1.9949(.089)[.085]	1.9950(.089)[.083]	1.9982(.072)	1.9977(.073)[.073]	1.9979(.073)[.071]
$\alpha_0 = -2$	-1.9943(.056)	-2.0001(.056)[.055]	-2.0002(.056)[.054]	-1.9908(.070)	-1.9977(.070)[.072]	-1.9981(.071)[.071]
$\tau_0 = -1$	-0.9712(.184)	-1.0091(.180)[.184]	-1.0085(.180)[.180]	-1.0490(.167)	-1.0263(.152)[.160]	-1.0251(.152)[.154]
$\sigma_0^2 = 1$	0.6054(.096)	0.9519(.150)[.136]	_	0.5960(.092)	0.9447(.146)[.137]	_
			n = 50	, T = 7		
$\beta_{10} = -1$	-0.9977(.056)	-0.9982(.056)[.054]	-0.9982(.056)[.054]	-1.0001(.049)	-1.0002(.049)[.046]	-1.0002(.049)[.046]
$\beta_{20} = 2$	1.9991(.051)	1.9994(.051)[.051]	1.9994(.051)[.050]	1.9980(.045)	1.9991(.045)[.045]	1.9992(.045)[.045]
$\alpha_0 = -2$	-1.9929(.034)	-2.0001(.034)[.036]	-2.0002(.034)[.035]	-1.9938(.039)	-2.0017(.039)[.039]	-2.0018(.039)[.039]
$\tau_0 = -1$	-0.9532(.107)	-0.9986(.105)[.103]	-0.9986(.105)[.103]	-0.9963(.083)	-1.0112(.081)[.080]	-1.0111(.081)[.080]
$\sigma_0^2 = 1$	0.8152(.067)	0.9845(.081)[.084]	_	0.8185(.067)	0.9855(.081)[.083]	_
			<i>n</i> = 100	0, T = 3		
$\beta_{10} = -1$	-1.0035(.069)	-1.0044(.069)[.068]	-1.0044(.069)[.067]	-1.0002(.060)	-1.0002(.060)[.061]	-1.0003(.060)[.060]
$\beta_{20} = 2$	1.9989(.063)	1.9983(.063)[.063]	1.9983(.063)[.062]	2.0009(.054)	2.0001(.054)[.054]	2.0001(.054)[.053]
$\alpha_0 = -2$	-1.9956(.042)	-1.9988(.042)[.043]	-1.9987(.042)[.042]	-1.9916(.053)	-1.9964(.053)[.052]	-1.9968(.053)[.051]
$\tau_0 = -1$	-1.0032(.131)	-0.9989(.127)[.130]	-0.9994(.128)[.131]	-1.0286(.102)	-1.0093(.096)[.101]	-1.0094(.096)[.099]
$o_{\bar{0}} = 1$	0.0205(.007)	0.9802(.107)[.103]	-	0.0305(.008)	0.9805(.100)[.101]	
	1.0000(007)	1 0000 ( 005) [ 005]	n = 100	$\frac{1}{2}, T = 7$		
$\beta_{10} = -1$	-1.0000(.037)	-1.0002(.037)[.037]	-1.0002(.037)[.037]	-0.9996(.033)	-0.9996(.033)[.033]	-0.9996(.033)[.033]
$\beta_{20} = 2$	1.9981(.038)	1.9982(.038)[.037]	1.9982(.038)[.037]	2.0015(.034)	2.0009(.034)[.034]	2.0009(.034)[.034]
$\alpha_0 = -2$	-1.9964(.024)	-1.9998(.024)[.024]	-1.9998(.024)[.024]	-1.9941(.028)	-1.9984(.029)[.028]	-1.9985(.029)[.028]
$\tau_0 = -1$ $\sigma_0^2 = 1$	-0.9811(.072) 0.8361(.051)	-1.0012(.072)[.070] 0.9939(.061)[.059]	-1.0012(.071)[.009]	-1.0109(.058) 0.8292(.050)	-1.0057(.050)[.057] 0.9923(.060)[.060]	-1.0035(.050)[.057]
			n = 50	T = 3		
$\beta = 1$	1.0026(.002)	1 0021( 002)[ 028]	1.0025(.002)[.086]	0.0064(.080)	0.0062(.088)[.087]	0.0061(.098)[.096]
$\beta_{10} = -1$ $\beta_{20} = 2$	-1.0020(.092) 2.0029(.094)	-1.0021(.092)[.088] 2 0010( 002)[ 080]	-1.0023(.092)[.080] 2 0014( 002)[ 086]	-0.9904(.089) 1 0062( 088)	-0.9902(.088)[.087] 1 9980( 087)[ 086]	-0.9901(.088)[.080] 1 0078( 087)[ 084]
$\rho_{20} = 2$ $\alpha_0 = -2$	-1.9914(.065)	-1.9977(.064)[.062]	-1.9974(.065)[.062]	-1.9891(.071)	-1.9946(.070)[.073]	-1.9949(.071)[.073]
$\tau_0 = 1$	1.2295(.259)	1.0731(.222)[.222]	1.0760(.227)[.222]	1.1229(.164)	1.0282(.151)[.156]	1.0284(.155)[.156]
$\sigma_0^2 = 1$	0.6000(.093)	0.9491(.147)[.138]		0.5950(.093)	0.9469(.148)[.138]	
n = 50, T = 7						
$\beta_{10} = -1$	-0.9972(.050)	-0.9973(.049)[.050]	-0.9973(.049)[.050]	-1.0019(.046)	-1.0013(.046)[.047]	-1.0013(.046)[.047]
$\beta_{20} = 2$	2.0005(.048)	1.9992(.048)[.045]	1.9992(.048)[.044]	2.0020(.044)	2.0010(.044)[.045]	2.0010(.044)[.045]
$\alpha_0 = -2$	-1.9953(.033)	-2.0014(.033)[.034]	-2.0014(.033)[.034]	-1.9906(.039)	-1.9977(.039)[.038]	-1.9977(.040)[.038]
$\tau_0 = 1$	1.1126(.121)	1.0203(.113)[.111]	1.0199(.113)[.110]	1.0590(.082)	1.0090(.080)[.080]	1.0090(.080)[.079]
$\sigma_{0}^{2} = 1$	0.8130(.071)	0.9838(.086)[.084]	—	0.8183(.071)	0.9865(.085)[.084]	—
			n = 100	), $T = 3$		
$\beta_{10} = -1$	-0.9996(.061)	-0.9989(.061)[.061]	-0.9991(.061)[.061]	-0.9992(.064)	-0.9990(.064)[.061]	-0.9989(.064)[.060]
$\beta_{20} = 2$	1.9963(.067)	1.9968(.067)[.064]	1.9969(.067)[.063]	1.9982(.057)	1.9990(.056)[.057]	1.9990(.056)[.057]
$\alpha_0 = -2$	-1.9970(.041)	-1.9996(.040)[.041]	-1.9994(.040)[.040]	-1.9893(.053)	-1.9945(.053)[.054]	-1.9942(.053)[.054]
$\tau_0 = 1$	1.1811(.177)	1.0320(.145)[.151]	1.0298(.146)[.151]	1.0713(.102)	1.0150(.096)[.101]	1.0144(.096)[.100]
$\sigma_0^2 = 1$	0.6126(.064)	0.9731(.101)[.103]	_	0.6289(.064)	0.9793(.099)[.101]	_
	n = 100, T = 7					
$\beta_{10} = -1$	-0.9976(.034)	-0.9978(.034)[.035]	-0.9978(.034)[.035]	-0.9995(.032)	-0.9999(.032)[.032]	-0.9999(.032)[.032]
$\beta_{20} = 2$	1.9988(.034)	1.9992(.034)[.034]	1.9992(.034)[.034]	2.0000(.036)	1.9996(.036)[.034]	1.9996(.036)[.034]
$\alpha_0 = -2$	-1.9972(.023)	-2.0002(.023)[.023]	-2.0002(.023)[.023]	-1.9953(.031)	-1.9997(.031)[.029]	-1.9996(.031)[.029]
$\tau_0 = 1$	1.0577(.075)	1.0124(.073)[.072]	1.0123(.073)[.072]	1.0444(.059)	1.0038(.057)[.057]	1.0041(.057)[.057]
$\sigma_{\tilde{0}} = 1$	0.8366(.051)	0.9951(.061)[.059]	_	0.8297(.051)	0.9939(.061)[.060]	—

Table 1: Homoske<br/>dastic case with  $\eta_{it} \sim N(0,1)$ 

	W=Rook, M=Queen			W=Queen, M=Rook		
	QMLE	ME	RME	QMLE	ME	RME
			n = 50	, T = 3		
$\beta_{10} = -1$	-1.0084(.094)	-1.0070(.094)[.093]	-1.0068(.094)[.090]	-0.9949(.086)	-0.9962(.086)[.082]	-0.9964(.086)[.082]
$\beta_{20} = 2$	1.9978(.089)	1.9966(.089)[.085]	1.9965(.089)[.085]	1.9988(.077)	1.9984(.077)[.073]	1.9985(.077)[.072]
$\alpha_0 = -2$	-1.9926(.053)	-1.9984(.053)[.055]	-1.9989(.053)[.053]	-1.9877(.071)	-1.9947(.071)[.072]	-1.9946(.072)[.069]
$\tau_0 = -1$	-0.9680(.182)	-1.0060(.179)[.184]	-1.0045(.179)[.180]	-1.0520(.168)	-1.0295(.153)[.158]	-1.0292(.153)[.151]
$\sigma_0^2 = 1$	0.6073(.094)	0.9549(.148)[.144]	_	0.6029(.095)	0.9556(.151)[.144]	_
			n = 50	, T = 7		
$\beta_{10} = -1$	-1.0004(.056)	-1.0009(.056)[.054]	-1.0009(.056)[.054]	-0.9984(.046)	-0.9985(.046)[.046]	-0.9985(.046)[.046]
$\beta_{20} = 2$	2.0028(.054)	2.0031(.054)[.051]	2.0031(.054)[.053]	1.9964(.046)	1.9976(.046)[.045]	1.9976(.046)[.046]
$\alpha_0 = -2$	-1.9950(.037)	-2.0022(.037)[.036]	-2.0021(.037)[.036]	-1.9890(.039)	-1.9969(.039)[.039]	-1.9971(.039)[.039]
$\tau_0 = -1$ $\sigma^2 = 1$	-0.9600(.104)	-1.0058(.103)[.104]	-1.0043(.103)[.102]	-0.9945(.078)	-1.0093(.075)[.080]	-1.0094(.075)[.079]
00 - 1	0.8102(.073)	0.9838(.088)[.090]	-	0.8185(.075)	0.9855(.090)[.090]	
	0.0000(.000)	0.0074(0.00)[0.00]	n = 100	1, T = 3	1 0010(000)[000]	
$\beta_{10} = -1$	-0.9962(.068)	-0.9971(.068)[.068]	-0.9972(.068)[.066]	-1.0013(.062)	-1.0013(.062)[.060]	-1.0013(.062)[.060]
$p_{20} = 2$	2.0018(.008) 1.0052(.044)	2.0012(.008)[.002] 1.0084(.044)[.042]	2.0012(.008)[.005] 1.0086(.044)[.042]	2.0031(.054) 1.0040(.052)	2.0024(.054)[.053] 1.0006(.052)[.051]	2.0024(.054)[.055] 1.0008(.052)[.052]
$\alpha_0 = -2$ $\tau_0 = -1$	-1.9953(.044) -1.0052(.136)	-1.9964(.044)[.043] -1.0000(.132)[.120]	-1.9980(.044)[.042] -1.0020(.133)[.130]	-1.9949(.032) -1.0325(.100)	-1.9990(.052)[.051] -1.0131(.104)[.100]	-1.9998(.055)[.052] -1.0145(.104)[.099]
$\sigma_0^2 = 1$	0.6194(.071)	-1.0009(.132)[.129] 0.9783(.111)[.109]	-1.0020(.100)[.100]	0.6273(.071)	0.9753(.110)[.106]	-1.0145(.104)[.035]
			n = 100	T = 7	0.0100(0)[00]	
$\beta_{10} = -1$	_1.0019(.037)	_1.0021(.037)[.037]	-1.0021(.037)[.037]	-0.9999(.032)	_0.0000(.033)[.033]	_0.0000(.032)[.033]
$\beta_{10} = -1$ $\beta_{20} = 2$	2,0004(,039)	2,0005(,039)[,037]	2,0005(,039)[,039]	2,0008(,036)	2,0002(,036)[,034]	$2\ 0002(\ 036)[\ 035]$
$\beta_{20} = 2$ $\alpha_0 = -2$	-1.9984(.023)	-2.0017(.023)[.024]	-2.0017(.023)[.024]	-1.9951(.029)	-1.9994(.029)[.028]	-1.9994(.029)[.029]
$\tau_0 = -1$	-0.9798(.073)	-0.9999(.072)[.070]	-0.9997(.072)[.069]	-1.0101(.056)	-1.0050(.054)[.057]	-1.0049(.054)[.057]
$\sigma_0^2 = 1$	0.8331(.053)	0.9904(.063)[.063]	_	0.8278(.053)	0.9907(.064)[.065]	_
			n = 50	, T = 3		
$\beta_{10} = -1$	-0.9982(.087)	-0.9986(.085)[.088]	-0.9989(.085)[.083]	-1.0056(.092)	-1.0056(.092)[.087]	-1.0056(.092)[.089]
$\beta_{20} = 2$	2.0043(.097)	2.0028(.097)[.090]	2.0032(.097)[.091]	1.9984(.091)	1.9997(.090)[.085]	2.0000(.091)[.086]
$\alpha_0 = -2$	-1.9929(.061)	-1.9990(.060)[.063]	-1.9992(.061)[.063]	-1.9902(.073)	-1.9958(.073)[.072]	-1.9969(.074)[.072]
$\tau_0 = 1$	1.2120(.250)	1.0575(.215)[.223]	1.0693(.226)[.226]	1.1289(.167)	1.0330(.153)[.157]	1.0328(.157)[.156]
$\sigma_0^2 = 1$	0.6020(.094)	0.9522(.148)[.148]	_	0.5938(.093)	0.9451(.148)[.142]	
			n = 50	, <i>T</i> = 7		
$\beta_{10} = -1$	-0.9989(.051)	-0.9987(.051)[.050]	-0.9987(.051)[.051]	-1.0013(.048)	-1.0005(.048)[.047]	-1.0005(.048)[.047]
$\beta_{20} = 2$	2.0020(.048)	2.0008(.048)[.045]	2.0008(.048)[.046]	2.0017(.049)	2.0006(.050)[.045]	2.0006(.050)[.047]
$\alpha_0 = -2$ $\sigma_0 = 1$	-1.9944(.034) 1 1105( 117)	-2.0005(.034)[.034] 1.0176(.110)[.112]	-2.0001(.034)[.034] 1.0160(.110)[.111]	-1.9915(.039) 1.0643(.082)	-1.9987(.039)[.038] 1.0130(.070)[.080]	-1.9989(.038)[.038] 1.0140(.070)[.080]
$\sigma_0^2 = 1$	0.8141(.077)	0.9851(.093)[.091]	-	0.8170(.032)	1.0139(.079)[.080] 0.9850(.089)[.089]	-
00-1	0.0111(.011)	0.0001(.000)[.001]	n - 100	T = 3	0.0000(.000)[.000]	
$\beta_{ii} = 1$	1.0031(.061)	1.0022(.060)[.061]	$\frac{n - 100}{1.0024(.060)[.060]}$	$\frac{1, 1-3}{1, 0012(062)}$	1.0010(.062)[.061]	1.0000(.062)[.050]
$\beta_{10} = -1$ $\beta_{20} = 2$	-1.0031(.001) 1 9971( 069)	-1.0022(.000)[.001] 1 9983( 069)[ 064]	-1.0024(.000)[.000] 1 9984( 069)[ 064]	-1.0012(.002) 1 9996( 058)	-1.0010(.002)[.001] 2 0002( 058)[ 057]	-1.0009(.002)[.059] 2 0003( 058)[ 057]
	-1.9970(.041)	-1.9996(.041)[.041]	-1.9992(.041)[.040]	-1.9923(.054)	-1.9974(.054)[.054]	-1.9978(.054)[.054]
$\tau_0 = 1$	1.1866(.186)	1.0359(.152)[.152]	1.0357(.154)[.153]	1.0768(.105)	1.0193(.099)[.102]	1.0176(.099)[.100]
$\sigma_0^2 = 1$	0.6121(.070)	0.9725(.111)[.108]	_	0.6254(.069)	0.9739(.108)[.106]	_
	n = 100, T = 7					
$\beta_{10} = -1$	-0.9986(.036)	-0.9988(.036)[.035]	-0.9989(.036)[.035]	-0.9991(.031)	-0.9995(.031)[.032]	-0.9995(.031)[.032]
$\beta_{20}=2$	1.9977(.037)	1.9981(.037)[.034]	1.9981(.037)[.036]	1.9984(.035)	1.9980(.035)[.034]	1.9980(.035)[.035]
$\alpha_0 = -2$	-1.9953(.023)	-1.9984(.023)[.023]	-1.9985(.023)[.023]	-1.9955(.029)	-1.9998(.029)[.029]	-1.9996(.029)[.029]
$\tau_0 = 1$	1.0534(.075)	1.0084(.072)[.072]	1.0082(.073)[.072]	1.0447(.058)	1.0041(.056)[.058]	1.0039(.056)[.057]
$\sigma_{0}^{2} = 1$	0.8333(.052)	0.9912(.062)[.063]	-	0.8266(.056)	0.9901(.067)[.064]	_

Table 2: Heteroske<br/>dastic case with  $\eta_{it} \sim N(0,1)$ 

	W=Rook, M=Queen				W=Queen, M=Rook	
	QMLE	ME	RME	QMLE	ME	RME
			n = 50	, T = 3		
$\beta_{10} = -1$	-0.9962(.094)	-0.9949(.094)[.093]	-0.9948(.094)[.090]	-0.9998(.077)	-0.9999(.077)[.073]	-0.9999(.077)[.071]
$\beta_{20} = 2$	2.0015(.088)	2.0003(.088)[.085]	2.0003(.087)[.082]	1.9958(.081)	1.9964(.081)[.076]	1.9965(.081)[.073]
$\alpha_0 = -2$	-1.9923(.053)	-1.9980(.053)[.055]	-1.9981(.053)[.054]	-1.9844(.079)	-1.9934(.079)[.077]	-1.9932(.079)[.075]
$\tau_0 = -1$	-0.9600(.186)	-0.9979(.182)[.183]	-0.9976(.182)[.177]	-1.0416(.157)	-1.0326(.146)[.154]	-1.0319(.148)[.145]
$\sigma_0^2 = 1$	0.6023(.126)	0.9470(.198)[.176]	-	0.5928(.132)	0.9393(.209)[.180]	-
			n = 50	, T = 7		
$\beta_{10} = -1$	-1.0004(.054)	-1.0013(.054)[.053]	-1.0013(.054)[.052]	-1.0010(.047)	-0.9999(.047)[.046]	-0.9998(.047)[.046]
$\beta_{20} = 2$	1.9981(.056)	1.9978(.056)[.054]	1.9978(.056)[.053]	1.9986(.046)	1.9985(.046)[.045]	1.9985(.046)[.044]
$\alpha_0 = -2$	-1.9931(.035)	-1.9998(.035)[.035]	-1.9998(.035)[.034]	-1.9940(.035)	-2.0014(.035)[.036]	-2.0014(.035)[.036]
$\tau_0 = -1$	-0.9554(.111)	-0.9992(.109)[.106]	-0.9990(.109)[.104]	-0.9957(.087)	-1.0083(.084)[.082]	-1.0084(.084)[.081]
$\sigma_0^2 = 1$	0.8172(.106)	0.9914(.128)[.122]	_	0.8185(.103)	0.9879(.124)[.122]	_
			n = 100	T = 3		
$\beta_{10} = -1$	-1.0001(.073)	-0.9994(.073)[.074]	-0.9994(.073)[.073]	-1.0011(.053)	-1.0022(.053)[.053]	-1.0022(.053)[.052]
$\beta_{20} = 2$	1.9988(.061)	1.9982(.061)[.062]	1.9982(.061)[.061]	1.9992(.056)	1.9999(.056)[.054]	1.9999(.056)[.054]
$\alpha_0 = -2$	-1.9967(.047)	-2.0007(.047)[.047]	-2.0007(.047)[.046]	-1.9932(.049)	-1.9974(.049)[.048]	-1.9975(.049)[.047]
$\tau_0 = -1$	-0.9954(.132)	-0.9996(.129)[.126] 0.0720(.144)[.127]	-0.9995(.129)[.123]	-1.0362(.103)	-1.0220(.099)[.100] 0.0742(.120)[.125]	-1.0212(.098)[.097]
00 - 1	0.0229(.092)	0.9739(.144)[.137]	-	0.0291(.090)	0.9745(.159)[.155]	_
	0.000.1(.000)		n = 100	T = 7	1 0000( 000)[ 001]	1 0000/ 000)[ 001]
$\beta_{10} = -1$	-0.9994(.039)	-0.9993(.039)[.039]	-0.9993(.039)[.038]	-1.0009(.032)	-1.0008(.032)[.031]	-1.0008(.032)[.031]
$p_{20} = 2$	1.9997(.039) 1.0055(.027)	2.0001(.039)[.038] 1.0002(.027)[.026]	2.0001(.039)[.038] 1.0002(.027)[.026]	1.9992(.031) 1.0055(.026)	1.9994(.031)[.030] 1.0001(.026)[.026]	1.9994(.031)[.030] 1.0001(.026)[.026]
$\alpha_0 = -2$	-1.9955(.027)	-1.9995(.027)[.020] 1.0002(.071)[.072]	-1.9995(.027)[.020] 1.0002(.071)[.071]	-1.9955(.020) 1.0102(.050)	-1.9991(.020)[.020] 1.0062(.057)[.057]	-1.9991(.020)[.020] 1.0062(.058)[.056]
$\sigma_0^2 = -1$ $\sigma_0^2 = 1$	-0.9839(.072) 0.8316(.075)	-1.0003(.071)[.072] 0.9935(.089)[.089]	-1.0003(.071)[.071]	-1.0105(.039) 0.8298(.076)	-1.0002(.037)[.037] 0.9915(.091)[.088]	-1.0002(.038)[.030]
- 0 -			n = 50	T = 3		
$\beta_{10} = -1$	_0.0007(.083)	_0.0081(.081)[.081]	_0.9986(.082)[.078]	$\frac{10055(096)}{-10055(096)}$	-1.0064(.096)[.095]	_1.0063(.006)[.003]
$\beta_{10} = -1$ $\beta_{20} = 2$	2.0014(.083)	$2\ 0014(\ 082)[\ 081]$	2.0015(.082)[.070]	1.0055(.050) 1.9958(.079)	-1.0004(.030)[.033] 1 9980( 078)[ 077]	(1.0003(.030)[.033] (1.9979(.078)[.074]
$\alpha_0 = -2$	-1.9913(.059)	-1.9967(.058)[.059]	-1.9972(.058)[.058]	-1.9836(.088)	-1.9958(.087)[.087]	-1.9955(.088)[.085]
$\tau_0 = 1$	1.2170(.250)	1.0721(.216)[.221]	1.0784(.230)[.223]	1.1200(.165)	1.0273(.151)[.159]	1.0262(.153)[.152]
$\sigma_0^2 = 1$	0.5997(.128)	0.9480(.203)[.180]	_	0.6000(.121)	0.9548(.192)[.186]	_
	n = 50, T = 7					
$\beta_{10} = -1$	-0.9997(.048)	-0.9998(.048)[.049]	-0.9998(.048)[.049]	-1.0007(.050)	-0.9999(.049)[.048]	-0.9999(.049)[.048]
$\beta_{20} = 2$	2.0001(.052)	2.0002(.052)[.052]	2.0002(.052)[.052]	1.9994(.051)	2.0010(.051)[.051]	2.0010(.051)[.051]
$\alpha_0 = -2$	-1.9932(.034)	-1.9987(.034)[.034]	-1.9987(.034)[.034]	-1.9906(.043)	-1.9997(.043)[.042]	-1.9996(.043)[.042]
$\tau_0 = 1$	1.1228(.125)	1.0211(.116)[.115]	1.0210(.116)[.114]	1.0644(.084)	1.0110(.081)[.081]	1.0108(.082)[.080]
$\sigma_{0}^{2} = 1$	0.8121(.107)	0.9877(.130)[.125]	-	0.8199(.115)	0.9911(.139)[.123]	_
			n = 100	T = 3		
$\beta_{10} = -1$	-0.9967(.067)	-0.9972(.067)[.066]	-0.9975(.067)[.065]	-0.9966(.055)	-0.9968(.055)[.056]	-0.9968(.055)[.055]
$\beta_{20} = 2$	1.9987(.064)	2.0003(.063)[.062]	2.0003(.063)[.061]	1.9995(.056)	1.9997(.056)[.058]	1.9997(.056)[.057]
$\alpha_0 = -2$	-1.9958(.048)	-1.9998(.047)[.046]	-1.9999(.048)[.046]	-1.9939(.046)	-1.9971(.046)[.045]	-1.9972(.046)[.045]
$\tau_0 = 1$	1.1221(.152)	1.0326(.139)[.141]	1.0353(.143)[.143]	1.0683(.099)	1.0176(.094)[.099]	1.0169(.094)[.097]
$\sigma_0^2 = 1$	0.6268(.093)	0.9824(.146)[.139]	—	0.6304(.094)	0.9776(.146)[.135]	—
			n = 100	), $T = 7$		
$\beta_{10} = -1$	-1.0003(.034)	-1.0005(.034)[.034]	-1.0005(.034)[.034]	-0.9998(.032)	-0.9997(.032)[.030]	-0.9997(.032)[.030]
$\beta_{20} = 2$	2.0007(.036)	2.0009(.036)[.035]	2.0010(.036)[.035]	2.0006(.036)	2.0011(.036)[.037]	2.0011(.036)[.037]
$\alpha_0 = -2$	-1.9972(.025)	-2.0005(.025)[.025]	-2.0005(.025)[.025]	-1.9971(.029)	-2.0016(.029)[.030]	-2.0016(.029)[.030]
$\tau_0 = 1$ $\sigma^2 = 1$	1.0663(.079)	1.0107(.076)[.075] 0.0011(.002)[.099]	1.0106(.075)[.075]	1.0402(.059) 0.8200(.074)	1.0038(.057)[.056] 0.0025(.020)[.020]	1.0038(.057)[.056]
00 - 1	0.0209(.018)	0.3311(.033)[.000]	-	0.0239(.014)	0.3320(.003)[.009]	—

Table 3: Homosked astic case with  $\eta_{it} \sim \text{Gamma}(2,1)$ 

	W=Rook, M=Queen				W=Queen, M=Rook	
	QMLE	ME	RME	QMLE	ME	RME
			n = 50	, T = 3		
$\beta_{10} = -1$	-1.0029(.091)	-1.0016(.090)[.093]	-1.0015(.090)[.091]	-1.0017(.078)	-1.0019(.078)[.074]	-1.0017(.078)[.073]
$\beta_{20} = 2$	2.0029(.091)	2.0017(.091)[.085]	2.0016(.091)[.084]	2.0001(.081)	2.0006(.081)[.077]	2.0006(.081)[.077]
$\alpha_0 = -2$	-1.9921(.053)	-1.9978(.053)[.054]	-1.9982(.053)[.053]	-1.9893(.079)	-1.9984(.079)[.078]	-1.9984(.079)[.076]
$\tau_0 = -1$	-0.9788(.181)	-1.0166(.177)[.185]	-1.0153(.177)[.177]	-1.0403(.152)	-1.0313(.142)[.155]	-1.0289(.142)[.147]
$\sigma_0^2 = 1$	0.6119(.133)	0.9621(.208)[.189]	-	0.6031(.134)	0.9557(.213)[.189]	-
			n = 50	, T = 7		
$\beta_{10} = -1$	-0.9981(.054)	-0.9990(.054)[.052]	-0.9990(.054)[.052]	-1.0009(.045)	-0.9997(.045)[.046]	-0.9997(.045)[.046]
$\beta_{20} = 2$	2.0018(.058)	2.0015(.058)[.053]	2.0015(.058)[.057]	2.0016(.046)	2.0015(.046)[.045]	2.0016(.046)[.047]
$\alpha_0 = -2$	-1.9937(.035)	-2.0003(.035)[.034]	-1.9999(.035)[.034]	-1.9922(.037)	-1.9996(.037)[.036]	-1.9994(.037)[.037]
$\tau_0 = -1$	-0.9579(.104)	-1.0021(.102)[.107]	-1.0020(.102)[.104]	-0.9952(.089)	-1.0077(.086)[.082]	-1.0080(.086)[.081]
$\sigma_0^2 = 1$	0.8056(.106)	0.9773(.129)[.129]	_	0.8181(.110)	0.9874(.132)[.130]	_
			n = 100	T = 3		
$\beta_{10} = -1$	-0.9977(.077)	-0.9970(.077)[.074]	-0.9970(.077)[.076]	-0.9988(.053)	-0.9998(.053)[.053]	-0.9999(.053)[.051]
$\beta_{20} = 2$	1.9972(.064)	1.9966(.064)[.062]	1.9966(.064)[.063]	1.9992(.057)	1.9999(.057)[.055]	1.9999(.057)[.056]
$\alpha_0 = -2$	-1.9922(.046)	-1.9962(.046)[.046]	-1.9962(.046)[.046]	-1.9964(.046)	-2.0007(.046)[.048]	-2.0008(.046)[.046]
$\tau_0 = -1$	-0.9971(.126)	-1.0012(.124)[.126] 0.0769(.152)[.142]	-1.0009(.124)[.124]	-1.0278(.102)	-1.0140(.097)[.100] 0.0780(.142)[.142]	-1.0131(.097)[.096]
00 - 1	0.0243(.098)	0.9702(.155)[.145]	-	0.0321(.093)	0.9789(.145)[.145]	_
1	1.0000(.000)	1 0000 ( 000) [ 000]	n = 100	D, T = 7	1 0014(001)[000]	1.0014(.001)[.001]
$\beta_{10} = -1$	-1.0008(.038)	-1.0008(.038)[.039]	-1.0008(.038)[.039]	-1.0015(.031)	-1.0014(.031)[.032]	-1.0014(.031)[.031]
$\beta_{20} = 2$	2.0013(.039)	2.0017(.039)[.038]	2.0017(.039)[.038]	2.0021(.031)	2.0022(.031)[.030]	2.0022(.031)[.031]
$\alpha_0 = -2$	-1.9904(.025)	-2.0003(.025)[.020] 1 0000( 071)[ 072]	-2.0004(.025)[.020] 1 0001( 071)[ 071]	-1.9959(.027) 1.0106(.057)	-1.9995(.027)[.026] 1.0065(.055)[.056]	-1.9990(.027)[.026] 1.0068(.055)[.056]
$\sigma_0^2 = 1$	-0.9830(.072) 0.8292(.081)	-1.0000(.071)[.072] 0.9906(.097)[.094]	-1.0001(.071)[.071]	-1.0100(.037) 0.8349(.081)	-1.0005(.055)[.050] 0.9976(.097)[.094]	-1.0008(.055)[.050]
			n = 50	T = 3		
$\beta_{10} = -1$	_1.0011(.085)	_0.0004(.084)[.081]	_0.0007(.084)[.078]	-0.9979(.093)	_0.0001(.002)[.004]	_0.0002(.002)[.088]
$\beta_{10} = -1$ $\beta_{20} = 2$	2.0009(.086)	2,0009(.086)[.082]	2,0014(.086)[.083]	1.9942(.086)	-0.9991(.092)[.094] 1 9961( 086)[ 076]	1.9960(.086)[.077]
$\alpha_0 = -2$	-1.9923(.058)	-1.9977(.058)[.059]	-1.9984(.058)[.057]	-1.9842(.088)	-1.9958(.087)[.086]	-1.9957(.087)[.085]
$\tau_0 = 1$	1.1957(.237)	1.0547(.207)[.218]	1.0643(.218)[.223]	1.1231(.161)	1.0316(.148)[.157]	1.0303(.149)[.150]
$\sigma_0^2 = 1$	0.6049(.133)	0.9559(.209)[.192]	_	0.5888(.128)	0.9369(.204)[.186]	_
	n = 50, T = 7					
$\beta_{10} = -1$	-1.0016(.050)	-1.0017(.050)[.049]	-1.0018(.050)[.048]	-1.0014(.047)	-1.0005(.047)[.048]	-1.0006(.047)[.048]
$\beta_{20} = 2$	2.0007(.053)	2.0010(.053)[.052]	2.0011(.053)[.054]	1.9969(.054)	1.9984(.054)[.051]	1.9984(.054)[.053]
$\alpha_0 = -2$	-1.9938(.034)	-1.9994(.034)[.034]	-1.9996(.034)[.034]	-1.9896(.041)	-1.9987(.041)[.042]	-1.9984(.041)[.041]
$\tau_0 = 1$	1.1247(.121)	1.0231(.113)[.114]	1.0236(.113)[.113]	1.0587(.085)	1.0057(.082)[.081]	1.0056(.082)[.080]
$\sigma_{0}^{2} = 1$	0.8133(.116)	0.9891(.141)[.130]	—	0.8173(.112)	0.9880(.136)[.131]	—
			n = 100	T = 3		
$\beta_{10} = -1$	-0.9964(.065)	-0.9971(.064)[.065]	-0.9973(.064)[.064]	-1.0001(.057)	-1.0001(.057)[.056]	-1.0001(.057)[.056]
$\beta_{20} = 2$	1.9980(.063)	1.9993(.063)[.062]	1.9994(.063)[.060]	1.9997(.061)	1.9999(.061)[.058]	2.0000(.061)[.060]
$\alpha_0 = -2$	-1.9964(.046)	-2.0002(.045)[.046]	-2.0004(.045)[.046]	-1.9955(.045)	-1.9987(.045)[.046]	-1.9990(.045)[.045]
$\tau_0 = 1$	1.1127(.157)	1.0249(.142)[.140]	1.0278(.144)[.141]	1.0637(.103)	1.0132(.098)[.099]	1.0129(.098)[.098]
$\sigma_0^2 = 1$	0.6180(.096)	0.9687(.151)[.142]	-	0.6290(.095)	0.9753(.147)[.141]	_
	n = 100, T = 7					
$\beta_{10} = -1$	-0.9993(.034)	-0.9996(.034)[.034]	-0.9996(.034)[.034]	-0.9998(.030)	-0.9997(.030)[.030]	-0.9997(.030)[.030]
$\beta_{20} = 2$	1.9988(.036)	1.9990(.036)[.035]	1.9990(.036)[.036]	1.9986(.038)	1.9991(.038)[.037]	1.9991(.038)[.038]
$\alpha_0 = -2$	-1.9961(.026)	-1.9994(.026)[.025]	-1.9994(.026)[.025]	-1.9942(.030)	-1.9986(.030)[.030]	-1.9986(.030)[.029]
$\tau_0 = 1$ $\sigma^2 = 1$	1.0679(.079)	1.0123(.076)[.075] 0.0804(.006)[.002]	1.0125(.076)[.075]	1.0441(.059) 0.8306(.079)	1.0070(.057)[.056]	1.0076(.057)[.056]
00 - 1	0.0214(.000)	0.3034(.030)[.035]	-	0.0300(.078)	0.33335(.033)[.094]	-

Table 4: Heteroskedastic case with  $\eta_{it} \sim \text{Gamma}(2, 1)$ 

#### 6 The third country effects on the US outward FDI stock

In this section, we consider an extension of the bilateral specification of the three-factor knowledge capital model suggested in Baltagi et al. (2007) to investigate the third country effects on the US outward FDI stock at the industry level. We collect data on the US outward FDI stock in 47 host countries at the industry level between 2008–2014. These host countries include developed, developing and undeveloped ones, and thus yielding a comprehensive sample.

In our analysis, each unit is a country-industry pair, denoted by a subscript *i*. The dependent variable is  $LFDI_{it}$ , which is the natural log of US outward FDI stock in country-industry pair i at year t, where  $i = 1, ..., n_t$  and t = 1, ..., T. In our sample, there are seven periods, i.e., T = 7, covering the period 2008–2014, and  $n_t$  varies from period to period, from a minimum of 340 countryindustry pairs in 2013 to a maximum of 364 country-industry pairs in 2011. The dataset is thus unbalanced with a total number of observations,  $N = \sum_{t=1}^{T} n_t = 2454$ , spanning 47 host countries and 10 industries. The set of independent variables includes a measure of absolute bilateral country size, given by  $G_{it} = \ln(\text{GDP}_{US,t} + \text{GDP}_{it})$ , where the US subscript denotes the parent country and the *i* subscript denotes the host country. We also include a similarity measure of country size proposed by Helpman and Krugman (1985) and Helpman (1987), and is given by  $S_{it} = 1 - s_{US,t}^2 - s_{it}^2$ where  $s_{US,t} = \text{GDP}_{US,t} / (\text{GDP}_{US,t} + \text{GDP}_{it})$  and  $s_{it} = \text{GDP}_{it} / (\text{GDP}_{US,t} + \text{GDP}_{it})$  are the share of respective country's GDP in the bilateral GDP. Another independent variable is a measure of relative capital stock defined by  $k_{it} = \ln(K_{US,t}/K_{it})$ , where  $K_{US,t}$  and  $K_{it}$  are the capital stocks of the US and the host country *i*, respectively. A measure of relative skilled labor endowment is also considered as an independent variable. This measure is defined by  $h_{it}^s = \ln(H_{US,t}/H_{it})$ , with  $H_{US,t}$  and  $H_{it}$  denoting the skilled labor endowments in the US and the host country i, respectively. Similarly, we also have a measure of relative unskilled labor endowment, given by  $h_{it}^{u} = \ln(L_{US,t}/L_{it})$ , with  $L_{US,t}$  and  $L_{it}$  denoting the unskilled labor endowments in the US and the host country *i*, respectively. The model includes two interaction terms:  $\Psi_{it} = G_{it} \times k_{it}$  and  $\Phi_{it} = \ln(d_{US,i} \times (k_{it} - h_{it}^u))$ , where  $d_{US,i}$  is the physical distance between the capitals of the US and the host country i. The last independent variable is an indicator of host country investment risk, denoted by R, computed as the inverse of an investment profile index. The sources and the descriptive statistics of our sample data are given in the web appendix.

	ME	RME
α	-0.0826	$-0.1692^{***}$
a	$(.057)^{-0.00}$	$(.053)^{-1}$
au	0.1201	0.2756***
	(.135)	(.063)
G	-4.0850	$-5.1272^{**}$
	(3.024)	(2.594)
S	$1.5493^{***}$	$1.5418^{**}$
,	(.455)	(.630)
k	-2.3859	4.5631
1.8	(10.734)	(9.262)
$n^{\circ}$	-0.0024	-0.4840
$h^u$	(.280) -1.0157	(.308) -1.2527
11	$(2\ 004)$	(1.871)
$\Psi$	0.1552	-0.1034
	(.372)	(.320)
$\Phi$	-0.2895	-0.1943
_	(.240)	(.212)
R	$0.0494^{***}$	$0.0447^{*}$
wa	(.019)	(.024)
WG	$-0.0292^{***}$	$-0.0340^{***}$
WC	(.010)	(.007)
WB	(260)	(175)
Wk	0.8501	0.9316
,, 10	(1.727)	(1.645)
$Wh^s$	0.1795*	0.2877***
	(.106)	(.081)
$Wh^u$	$1.2566^{***}$	$1.3541^{***}$
	(.481)	(.414)
$W\Phi$	-0.0398	-0.0447
TL/ST/	(.055)	(.054)
$W \Psi$	$(0.1139^{**})$	$(0.1192^{***})$
WD	(.001)	(.040)
WN	(0.0252)	$(0.0200)^{-0.0200}$
$\sigma^2$	0.2508***	()
0	(0.2598)	
$R^2$	94.93%	94.93%

Table 5: Estimation results for the US outward FDI Stock

Significance levels: \*: 10%, \*\*: 5%, and \*\*\*: 1%.

We augment the model suggested in Baltagi et al. (2007) with three types of spatial dependence: (i) spatial dependence in FDI specified through a matrix exponential term, (ii) spatial lags of the third-country determinants of FDI, and (iii) spatial dependence in the error terms specified through a matrix exponential term. Our estimation equation allows for heterogeneity in two dimensions, namely, the country-industry pairs and time dimensions. Thus, our estimation equation takes the following form:

$$e^{\alpha W_{t}} LFDI_{t} = \beta_{0}l_{n_{t}} + \beta_{1}G_{t} + \beta_{2}S_{t} + \beta_{3}k_{t} + \beta_{4}h_{t}^{s} + \beta_{5}h_{t}^{u} + \beta_{6}\Psi_{t} + \beta_{7}\Phi_{t} + \beta_{8}R_{t} + \beta_{9}W_{t}G_{t} + \beta_{10}W_{t}S_{t} + \beta_{11}W_{t}k_{t} + \beta_{12}W_{t}h_{t}^{s} + \beta_{13}W_{t}h_{t}^{u} + \beta_{14}W_{t}\Psi_{t} + \beta_{15}W_{t}\Phi_{t} + \beta_{16}W_{t}R_{t} + C_{t}\mu + \lambda_{t}l_{n_{t}} + u_{t}, \quad e^{\tau W_{t}}u_{t} = \epsilon_{t}.$$

The elements of the  $n_t \times n_t$  spatial weights matrix  $W_t$  are specified as  $w_{ij} = d_{ij}^{-1} / \sum_{j=1}^N d_{ij}^{-1}$  for  $i \neq j$ and  $w_{ij} = 0$  for i = j, and  $d_{ij}$  is the bilateral physical distance between capitals of i and j taken from Mayer and Zignago (2011). The country-industry pairs and time fixed effects are respectively denoted by  $\mu$  and  $\lambda_t$ . We also allow for the presence of an unknown form of heteroskedasticity in the error term  $\epsilon_t$ .

We use our suggested M-estimators to estimate the parameters of the model. Table 5 provides the estimation results. The first column contains the estimation results obtained under homoskedasticity assumption from the ME. The second column provides the estimation results obtained from the RME under the assumption of an unknown form of heteroskedasticity. In the case of spatial parameters, the ME reports statistically insignificant results, whereas the RME reports statistically significant results. The RME estimates of  $\alpha$  and  $\tau$  are -0.1692 and 0.2756, respectively. Since the US outward FDI stock is likely to be affected differently by the unobserved factors in different host countries, it is more likely that the disturbance terms are heteroskedastic in our model. The ME reports a negative and insignificant estimate for G, while the RME reports a negative and statistically significant estimate. Both estimators report a positive and statistically significant estimate for the similarity variable, which is in line with prediction of the bilateral specification of the three-factor knowledge capital model suggested in Baltagi et al. (2007). In the case of k,  $h^u$ ,  $\Psi$  and  $\Phi$ , both estimators report statistically insignificant estimates.

The third-country effects are captured by the spatially weighted regressors. We can see that most of their coefficient estimates are significant, which means that the third-country effects are important and should not be neglected. Both estimators report negative and statistically significant estimate for the spatial lag of bilateral country size. This result indicates that there can be reallocation of plants from the host country to the third country when there is an increase in the size of third country (Baltagi et al., 2007). Both estimators report positive and statistically significant estimates for the spatially weighted similarity variable and the relative skilled labor endowment variable. These results are in line with the horizontal and export-platform FDI modes described in Baltagi et al. (2007). Both estimators report positive and statistically significant estimates for the relative unskilled labor endowment, which supports horizontal and vertical FDI modes. Both estimators report negative and statistically insignificant estimates for  $W\Phi$ , while positive and significant estimates for  $W\Psi$ . These results support vertical mode in the case of  $W\Phi$ , and horizontal, export-platform and complex vertical modes in the case of  $W\Psi$ . The last row of the table provides a pseudo- $R^2$  measure for model fit. Both estimators report a pseudo- $R^2$  value of 94.93%. Finally, both estimators report negative and statistically significant estimates for WR. Overall, our results show that the third-country effects are important in the US outward FDI stock.

## 7 Conclusion

In this paper, we considered a direct likelihood based estimation of the matrix exponential unbalanced panel data model that has the entity and time fixed effects. Our estimation approach allowed for the estimation of the common parameters and fixed effects simultaneously. We first showed that the QMLE of our model may not be consistent under both homoskedastic and heteroskedastic disturbance terms. We then adjusted the score functions under both homoskedastic and heteroskedastic cases, and defined valid M-estimators based on the adjusted score functions. We established the large sample properties of the resulting M-estimators, the ME and the RME, respectively. For inference, we suggested an analytical bias correction approach involving the sample counterpart and plug-in methods to consistently estimate the variance-covariance matrices of the suggested estimators. Through an extensive Monte Carlo study, we showed that our suggested estimators have good finite sample properties. In an empirical application, we showed that there is statistical evidence for the presence of third country effects on the US outward FDI stock.

## Supplementary materials

Additional materials are given in an accompanying web appendix.

## Acknowledgement

We are grateful to the editor, the associate editor and two anonymous referees for helpful comments on an earlier version of this paper.

## Funding

Ye Yang gratefully acknowledges the financial support from the Special Research Fund of Beijing for

Capital University of Economics and Business (ZD202104) and the research fund for new professors (XRZ2023042).

## **Conflict of Interest**

The authors report there are no competing interests to declare.

## References

- Antweiler, Werner (2001). "Nested random effects estimation in unbalanced panel data". In: Journal of Econometrics 101.2, pp. 295–313.
- Bai, Jushan, Yuan Liao, and Jisheng Yang (2015). "Unbalanced panel data models with interactive effects". In: *The Oxford Handbook of Panel Data*, pp. 149–170.
- Baltagi, Badi H. and Young-Jae Chang (1994). "Incomplete panels: A comparative study of alternative estimators for the unbalanced one-way error component regression model". In: *Journal of Econometrics* 62.2, pp. 67–89.
- Baltagi, Badi H., Seuck H. Song, and Byoung C. Jung (2001). "The unbalanced nested error component regression model". In: *Journal of Econometrics* 101.2, pp. 357–381.
- Baltagi, Badi H, Peter Egger, and Michael Pfaffermayr (2007). "Estimating models of complex FDI: Are there thirdcountry effects?" In: *Journal of Econometrics* 140.1, pp. 260–281.
- Blonigen, Bruce A, Ronald B Davies, Glen R Waddell, and Helen T Naughton (2007). "FDI in space: Spatial autoregressive relationships in foreign direct investment". In: *European Economic Review* 51.5, pp. 1303–1325.
- 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.
- Coughlin, Cletus C. and Eran Segev (2000). "Foreign Direct Investment in China: A Spatial Econometric Study". In: The World Economy 23.1, pp. 1–23.
- Cressie, Noel A.C. (1993). Statistics for Spatial Data. Revised Edition. New York: John Wiley & Sons, Inc.
- Davis, Peter (2002). "Estimating multi-way error components models with unbalanced data structures". In: *Journal* of *Econometrics* 106.1, pp. 67–95.
- 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, pp. 1–21.
- Helpman, Elhanan (1987). "Imperfect competition and international trade: Evidence from fourteen industrial countries". In: Journal of the Japanese and international economies 1.1, pp. 62–81.
- Helpman, Elhanan and Paul R Krugman (1985). Market structure and foreign trade: Increasing returns, imperfect competition, and the international economy. MIT press.
- 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 (2004). "Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models". In: *Econometrica* 72.6, pp. 1899–1925.
- 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.
- Leonard, Tom and John S. J. Hsu (1992). "Bayesian Inference for a Covariance Matrix". In: *The Annals of Statistics* 20.4, pp. 1669–1696.
- 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.
- Li, Liyao and Zhenlin Yang (2021). "Spatial dynamic panel data models with correlated random effects". In: *Journal of Econometrics* 221.2, pp. 424–454.
- Lin, Xu and Lung-fei Lee (2010). "GMM estimation of spatial autoregressive models with unknown heteroskedasticity". In: Journal of Econometrics 157, pp. 34–52.
- Liu, Shew Fan and Zhenlin Yang (2015). "Modified QML estimation of spatial autoregressive models with unknown heteroskedasticity and nonnormality". In: *Regional Science and Urban Economics* 52, pp. 50–70.
- Mayer, Thierry and Soledad Zignago (2011). Notes on CEPII's distances measures: The GeoDist database. Working Papers 2011-25. CEPII.
- Meng, Xiaoyu and Zhenlin Yang (2021). Unbalanced Spatial Panel Data Models with Fixed Effects. Tech. rep. Working paper, School of Economics, Singapore Management University.
- Neyman, J. and Elizabeth L. Scott (1948). "Consistent Estimates Based on Partially Consistent Observations". In: *Econometrica* 16.1, pp. 1–32.
- van der Vaart, A. W. (1998). Asymptotic Statistics. Cambridge UK: Cambridge university press.
- Wang, Wei and Lung fei Lee (2013). "Estimation of spatial panel data models with randomly missing data in the dependent variable". In: *Regional Science and Urban Economics* 43.3, pp. 521–538.
- Wansbeek, Tom and Arie Kapteyn (1989). "Estimation of the error-components model with incomplete panels". In: Journal of Econometrics 41.3, pp. 341–361.
- Wooldridge, Jeffrey M. (2019). "Correlated random effects models with unbalanced panels". In: Journal of Econometrics 211.1, pp. 137–150.
- 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.
- Yang, Zhenlin (2018). "Unified M-estimation of fixed-effects spatial dynamic models with short panels". In: Journal of Econometrics 205.2, pp. 423–447.
- Zhang, Yuanqing, Shuhui Feng, and Fei Jin (2019). "QML estimation of the matrix exponential spatial specification panel data model with fixed effects and heteroskedasticity". In: *Economics Letters* 180, pp. 1–5.