Accepted by Econometric Reviews on February 10th, 2022.

Model Selection and Model Averaging for Matrix Exponential Spatial Models

Ye Yang* Osman Doğan† Süleyman Taşpınar‡

February 11th, 2022

Abstract

In this paper, we focus on a model specification problem in spatial econometric models when an empiricist needs to choose from a pool of candidates for the spatial weights matrix. We propose a model selection (MS) procedure for the matrix exponential spatial specification (MESS), when the true spatial weights matrix may not be in the set of candidate spatial weights matrices. We show that the selection estimator is asymptotically optimal in the sense that asymptotically it is as efficient as the infeasible estimator that uses the best candidate spatial weights matrix. The proposed selection procedure is also consistent in the sense that when the data generating process involves spatial effects, it chooses the true spatial weights matrix with probability approaching one in large samples. We also propose a model averaging (MA) estimator that compromises across a set of candidate models. We show that it is asymptotically optimal. We further flesh out how to extend the proposed selection and averaging schemes to higher order specifications and to the MESS with heteroskedasticity. Our Monte Carlo simulation results indicate that the MS and MA estimators perform well in finite samples. We also illustrate the usefulness of the proposed MS and MA schemes in a spatially augmented economic growth model.

JEL-Classification: C13, C21, C52.
Keywords: Matrix Exponential Spatial Models, MESS, Model Selection, Model Averaging, Asymptotic Optimality, AMSE.

*School of Accounting, Capital University of Economics and Business, Beijing, China, email: yang.ye@cueb.edu.cn
†Department of Economics, University of Illinois at Urbana-Champaign, Illinois, U.S.A., email: odogan@illinois.edu
‡Department of Economics, Queens College, The City University of New York, New York, U.S.A., email: staspinar@qc.cuny.edu

1 Introduction

Spatial econometric models have been widely used (i) to model spillover effects in economic growth, international trade and foreign direct investment (Baltagi et al., 2007, 2008, 2016, 2017; Behrens et al., 2012; Desmet and Rossi-Hansberg, 2009; Ertur and Koch, 2007, 2011; König et al., 2019; Lee and Yu, 2012), (ii) to model peer effects, social interactions and production networks (Acemoglu et al., 2012; Bramoullé et al., 2009; Calvo-Armengol et al., 2009; Hsieh and Lee, 2016; Lee et al., 2020; Lin, 2010; Patacchini and Zenou, 2016; Pesaran and Yang, 2020), and (iii) to model spatial externalities (Bailey et al., 2016; Case, 1991; Kelejian and Piras, 2014; Kelejian et al., 2013). One main problem an empiricist faces in spatial econometric models is how to specify the spatial weights matrix (or the connectivity matrix) from a pool of candidates. Often it is the case that the spatial weights matrix is specified in an ad-hoc manner, and an extensive sensitivity analysis is carried out to justify the choice (Corrado and Fingleton, 2012).

In this paper, we focus on a model specification problem in terms of choosing a spatial weights matrix from a pool of candidates for the matrix exponential spatial specification (MESS). We propose a model selection (MS) procedure and show that the selection estimator is asymptotically optimal in the sense that it is as efficient as the infeasible estimator that uses the best candidate spatial weights matrix. Also, the proposed selection procedure chooses the true spatial weights matrix with probability approaching one in large samples when the data generating process involves spatial effects. We also propose a model averaging (MA) estimator that compromises across a set of candidate models and show that it also retains the asymptotic optimality.

The spatial econometric literature on model specification has mainly focused on testing the existence of spatial effects (Anselin, 1988; Anselin et al., 1996; Baltagi and Li, 2001; Baltagi and Liu, 2016; Baltagi and Yang, 2013; Born and Breitung, 2011; Doğan et al., 2018; Kelejian and Prucha, 2001; Lee and Yu, 2012; Taşpınar et al., 2018). In terms of choosing across a set of non-nested spatial effects, Anselin (1986) and Kelejian (2008) propose a J-test for spatial autoregressive (SAR) specifications. Here, the alternative models involve alternative weights matrices and the J-test utilizes whether or not the alternatives add to the explanatory power of the null model. Several improvements have been suggested to the J-test (Burridge, 2012; Burridge and Fingleton, 2010; Kelejian and Piras, 2011). The main shortcoming of the J-test is the fact that rejecting the null model based on the J-test does not suggest a formal way of choosing between the alternatives. Jin and Lee (2013) propose Cox-type tests of non-nested hypotheses for spatial autoregressive specifications and show that the Cox-type and J-type tests for non-nested hypotheses are not asymptotically equivalent under the null hypothesis. Similar to the J-test, there is no formal way of choosing amongst a set of alternatives when the Cox-type test rejects the null hypothesis. Han and Lee (2013) propose a J-test to choose between the MESS and SAR specification. The MESS and SAR models are non-nested and imply differing rates of decay for the spatial correlations.

In spatial econometrics, the MS problem in terms of choosing among a set of candidate weights matrices recently received some attention from the frequentist econometrics side.[1] Zhang and Yu

[1] In the spatial econometric literature, there are also some studies using the Bayesian methods for the model

(2018) formally study the MS problem in terms of choosing a weights matrix for the SAR model. They propose a MS procedure based on a Mallows type criterion function (Mallows, 1973), and show that their MS procedure is asymptotically optimal, and it is consistent when the true weights matrix is in the set of candidates.[2] They also propose a MA estimator and formally establish its asymptotic optimality under certain conditions.

Our focus in this study, on the other hand, is the MS and MA procedures for the MESS models, and specifically MESS$(1,1)$ and MESS(p,q). Originally, the MESS was suggested by LeSage and Pace (2007) as an alternative to the SAR type models because the likelihood estimation is greatly simplified as its likelihood function does not involve any Jacobian terms. Furthermore, there is no need to impose restrictions on the parameter space of the spatial parameters in the MESS, i.e., a MESS model always has a reduced form. The formal results for the maximum likelihood (ML) and generalized method of moments (GMM) estimation of the MESS models are established in Debarsy et al. (2015).

These attractive properties of the MESS models also make it more suitable for studying the MS and MA problems relative to the SAR type models. First, since the likelihood functions of the MESS type models do not involve any Jacobian terms, closed form expressions are available for some complicated terms in the Mallows type criterion function that we propose. This feature also allows us to extend our MS and MA procedures to higher order MESS models. Second, when the model involves heteroskedasticity of an unknown form in the error terms, our analysis need not be altered greatly as the (quasi) ML estimator remains consistent given that the spatial weights matrices are commutative (Debarsy et al., 2015). When the spatial weights matrices are not commutative, we suggest a robust GMM approach to form the Mallows type MS/MA criterion function. Despite these advantages of the MESS type models, it is important to note that the formal analysis is more difficult and complicated for the spatial models as opposed to the linear regression models, because the MS and MA procedures involve the ML/GMM estimators that are nonlinear functions of the outcome variable.

The remainder of this paper is organized as follows. Section 2 presents the model under consideration and lays out the details on the likelihood based estimator along with some terms needed in the construction of our suggested criterion function. Section 3 provides the details on the MS procedure and formally establishes the asymptotic optimality and the consistency of the selection estimator. Section 4 introduces our suggested MA procedure and establishes its asymptotic optimality. Section 5 shows how the proposed methods can be extended to higher order specifications, and to the MESS with heteroskedastic error terms. Section 6 presents the setting for our simulation study and the results. Section 7 shows how the proposed MS and MA schemes can be applied in modeling spillover effects in a model of economic growth. We conclude in Section 8 with some directions for future research. All technical details are collected in an appendix.

selection and model averaging procedures. Among others, see Debarsy and LeSage (2020), Han et al. (2017), LeSage and Pace (2009), and LeSage (2014).

[2]See Zhang and Yu (2018, p.2) for a discussion on the problems associated with the commonly used information criteria AIC and BIC in choosing amongst a set of weights matrices in spatial models.

2 The matrix exponential specification

We consider the following cross-sectional MESS$(1,1)$ model

$$e^{\alpha W} y = X\beta + u, \quad e^{\tau M} u = \epsilon, \tag{2.1}$$

where $y = (y_1, \ldots, y_n)^{'}$ is the $n \times 1$ vector of an outcome variable, X is the $n \times k$ matrix of non-stochastic exogenous variables with the matching parameter vector β, W and M are the $n \times n$ spatial weights matrices of known constants with zero diagonal elements. We refer $u = (u_1, \ldots, u_n)^{'}$ as the $n \times 1$ vector of regression error terms and $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^{'}$ as the $n \times 1$ vector of idiosyncratic errors (or innovations). We assume that ϵ_i's are independent and identically distributed (i.i.d.) with mean zero and variance σ^2. The scalar parameters α and τ are called the spatial coefficients that capture the intensity of interactions between observations. Let $C^0 = I_n$ for any an $n \times n$ matrix C, where I_n is the $n \times n$ identity matrix. Then, the matrix exponential function in (2.1) is defined as $e^{\alpha C} = \sum_{i=0}^{\infty} \frac{\alpha^i C^i}{i!}$ and is always invertible with inverse $e^{-\alpha C}$ (Chiu et al., 1996). Therefore, the reduced form of the MESS(1,1) always exists and is given by

$$y = e^{-\alpha W} X\beta + e^{-\alpha W} e^{-\tau M} \epsilon = \mu + \tilde{\epsilon}, \tag{2.2}$$

where $\mu = \mathbb{E}(y) = e^{-\alpha W} X\beta$ and $\tilde{\epsilon} = e^{-\alpha W} e^{-\tau M} \epsilon$.

We are interested in selecting a tuple of spatial weights matrices (W_s, M_s) from a set of candidate weights matrices, $\mathcal{W} = \{(W_s, M_s) : s \in \{1, 2, \ldots, S\}\}$, where we allow S to increase as the sample size n increases.[3] The quasi log-likelihood function of (2.1) based on the tuple (W_s, M_s) can be expressed as

$$\ell_s = -\frac{n}{2} \ln 2\pi\sigma^2 - \frac{1}{2\sigma^2} \left\| e^{\tau M_s} (e^{\alpha W_s} y - X\beta) \right\|^2, \tag{2.3}$$

where $\|\cdot\|$ denotes the Euclidean norm. Note that (2.3) does not include the Jacobian terms, because $\ln |e^{\alpha W_s}| = \ln \left(e^{\alpha \mathrm{tr}(W_s)}\right) = 0$ and $\ln |e^{\tau M_s}| = \ln \left(e^{\tau \mathrm{tr}(M_s)}\right) = 0$, where $|\cdot|$ denotes the determinant operator, and $\mathrm{tr}(\cdot)$ is the trace operator. This is one of the attractive properties of the MESS over the SAR type models, because the maximization of (2.3) does not involve the computation of the $n \times n$ log-Jacobian terms.[4] As we will show, the simpler form of the MESS quasi log-likelihood function also simplifies the derivation of some terms in our suggested selection criterion. For a

[3] Note that $\mathcal{W}$ includes all combinations. For example, if we had two candidate weights matrices A and B, then the subscript $s = 1$ could correspond to (A, A), $s = 2$ to (A, B), $s = 3$ to (B, A), and $s = 4$ to (B, B).

[4] Another advantage is that the MESS has an unrestricted parameter space for the spatial coefficients α and τ. On the other hand, for the SAR type models, the parameter space for the spatial coefficients have to be restricted for stability. See Kelejian and Prucha (2010) on the parameter space specifications for the SAR type models.

given $(\widehat{\alpha}_s, \widehat{\tau}_s)$ value, the first order conditions of (2.3) with respect to β and σ^2 yield

$$\widehat{\beta}_s = \left(X' e^{\widehat{\tau}_s M_s'} e^{\widehat{\tau}_s M_s} X\right)^{-1} X' e^{\widehat{\tau}_s M_s'} e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s} y, \tag{2.4}$$

$$\widehat{\sigma}_s^2 = \frac{1}{n} \left\| e^{\widehat{\tau}_s M_s} (e^{\widehat{\alpha}_s W_s} y - X\widehat{\beta}_s) \right\|^2. \tag{2.5}$$

Substituting (2.4) into $\widehat{\mu}_s = e^{-\widehat{\alpha}_s W_s} X \widehat{\beta}_s$, we obtain

$$\begin{aligned}\widehat{\mu}_s &= e^{-\widehat{\alpha}_s W_s} X \left(X' e^{\widehat{\tau}_s M_s'} e^{\widehat{\tau}_s M_s} X\right)^{-1} X' e^{\widehat{\tau}_s M_s'} e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s} y = e^{-\widehat{\alpha}_s W_s} e^{-\widehat{\tau}_s M_s} \widehat{P}_s e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s} y \\ &= \widetilde{P}_s y, \end{aligned} \tag{2.6}$$

where $\widehat{P}_s = e^{\widehat{\tau}_s M_s} X \left(X' e^{\widehat{\tau}_s M_s'} e^{\widehat{\tau}_s M_s} X\right)^{-1} X' e^{\widehat{\tau}_s M_s'}$ and $\widetilde{P}_s = e^{-\widehat{\alpha}_s W_s} e^{-\widehat{\tau}_s M_s} \widehat{P}_s e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s}$. We will use $\widehat{\mu}_s$ in constructing the loss function in the next section, which is the origin of the selection criterion for the spatial weights matrices.

3 Spatial weights matrix selection

Our goal is to select the spatial weights matrices that minimize a certain criterion function. As such, we follow the literature on the MS and MA problems (Hansen, 2007; Li, 1987; Wan et al., 2010; Zhang and Yu, 2018) to determine a selection criterion function in our setting. Let $L_s = \|\widehat{\mu}_s - \mu\|^2$ be the squared loss function, and let $R_s = \mathbb{E}\|\widehat{\mu}_s - \mu\|^2$ be the associated risk function. Using $\mu = y - \tilde{\epsilon}$, we can express R_s as

$$\begin{aligned} R_s &= \mathbb{E}\left\|\widetilde{P}_s y - y\right\|^2 + 2\mathbb{E}\left(\left(\widetilde{P}_s y - y\right)' \tilde{\epsilon}\right) + \mathbb{E}\left(\tilde{\epsilon}' \tilde{\epsilon}\right) \\ &= \mathbb{E}\left\|\widetilde{P}_s y - y\right\|^2 + 2\mathbb{E}\left(\left(\widetilde{P}_s y\right)' (y - \mu)\right) - \operatorname{tr}(\Omega), \end{aligned} \tag{3.1}$$

where $\Omega = \sigma^2 e^{-\alpha W} e^{-\tau M} e^{-\tau M'} e^{-\alpha W'}$ is the variance of y. We will select the spatial weights matrices by minimizing a modified version of R_s. To motivate the selection criterion function, assume further that ϵ_i's are normally distributed, and let $z = \Omega^{-\frac{1}{2}}(y - \mu)$. Then, we have

$$\begin{aligned} \mathbb{E}\left(\left(\widetilde{P}_s y\right)' (y - \mu)\right) &= \mathbb{E}\left(\left(\widetilde{P}_s \Omega^{\frac{1}{2}} z + \widetilde{P}_s \mu\right)' \Omega^{\frac{1}{2}} z\right) = \mathbb{E}\left(\operatorname{tr}\left(\frac{\partial\left(\Omega^{\frac{1}{2}} \widetilde{P}_s \Omega^{\frac{1}{2}} z\right)}{\partial z'}\right)\right) \\ &= \mathbb{E}\left(\operatorname{tr}\left(\Omega^{\frac{1}{2}} \widetilde{P}_s \Omega^{\frac{1}{2}}\right) + \operatorname{tr}\left(\frac{\partial \Omega^{\frac{1}{2}} \widetilde{P}_s \Omega^{\frac{1}{2}} z}{\partial \widehat{\alpha}_s} \frac{\partial \widehat{\alpha}_s}{\partial z'}\right) + \operatorname{tr}\left(\frac{\partial \Omega^{\frac{1}{2}} \widetilde{P}_s \Omega^{\frac{1}{2}} z}{\partial \widehat{\tau}_s} \frac{\partial \widehat{\tau}_s}{\partial z'}\right)\right) \\ &= \mathbb{E}\left(\operatorname{tr}\left(\widetilde{P}_s \Omega\right) + \frac{\partial \widehat{\alpha}_s}{\partial y'} \Omega \frac{\partial \widetilde{P}_s}{\partial \widehat{\alpha}_s} y + \frac{\partial \widehat{\tau}_s}{\partial y'} \Omega \frac{\partial \widetilde{P}_s}{\partial \widehat{\tau}_s} y\right), \end{aligned} \tag{3.2}$$

where the second equality follows from Lemma A.1 in Appendix A.[5] Based on (3.1) and (3.2), we propose the following Mallows C_p type selection criterion function,

$$C_s = \left\| \widetilde{P}_s y - y \right\|^2 + 2 \left(\mathrm{tr}\left(\widetilde{P}_s \Omega \right) + \frac{\partial \widehat{\alpha}_s}{\partial y'} \Omega \frac{\partial \widetilde{P}_s}{\partial \widehat{\alpha}_s} y + \frac{\partial \widehat{\tau}_s}{\partial y'} \Omega \frac{\partial \widetilde{P}_s}{\partial \widehat{\tau}_s} y \right), \tag{3.3}$$

such that $E(C_s) = R_s + \mathrm{tr}(\Omega)$, where we need to determine $\frac{\partial \widetilde{P}_s}{\partial \widehat{\alpha}_s}$, $\frac{\partial \widetilde{P}_s}{\partial \widehat{\tau}_s}$, $\frac{\partial \widehat{\alpha}_s}{\partial y}$ and $\frac{\partial \widehat{\tau}_s}{\partial y}$. Then, from the definition of $\widetilde{P}_s$, it follows that

$$\frac{\partial \widetilde{P}_s}{\partial \widehat{\alpha}_s} = -W_s \widetilde{P}_s + \widetilde{P}_s W_s, \tag{3.4}$$

$$\frac{\partial \widetilde{P}_s}{\partial \widehat{\tau}_s} = e^{-\widehat{\alpha}_s W_s} e^{-\widehat{\tau}_s M_s} \widehat{P}_s (M_s + M_s') \widehat{G}_s e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s}, \tag{3.5}$$

where $\widehat{G}_s = I_n - \widehat{P}_s$ and I_n is the $n \times n$ identity matrix. By using the implicit differentiation on $\frac{\partial \ell_s^c}{\partial \widehat{\alpha}_s}$ and $\frac{\partial \ell_s^c}{\partial \widehat{\tau}_s}$, where l_s^c is the concentrated log-likelihood function associated with (2.3), we can determine $\frac{\partial \widehat{\alpha}_s}{\partial y}$ and $\frac{\partial \widehat{\tau}_s}{\partial y}$. Since the log-likelihood function of the MESS does not involve the Jacobian terms, the closed-forms for these expressions can be conveniently obtained. As shown in Appendix B, we have

$$\frac{\partial \widehat{\alpha}_s}{\partial y} = -\frac{2a_1 A_1 - \left\| \widehat{G}_s e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s} y \right\|^2 (A_2 + A_3)}{2a_1^2 - \| \widehat{G}_s e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s} y \|^2 (a_2 + a_3)}, \tag{3.6}$$

$$\frac{\partial \widehat{\tau}_s}{\partial y} = -\frac{2b_1 B_1 - 2 \left\| \widehat{G}_s e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s} y \right\|^2 B_2}{b_1^2 - \left\| \widehat{G}_s e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s} y \right\|^2 (2b_2 + b_3)}, \tag{3.7}$$

where

$$a_1 = y' e^{\widehat{\alpha}_s W_s'} e^{\widehat{\tau}_s M_s'} \widehat{G}_s e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s} W_s y, \quad a_2 = y' W_s' e^{\widehat{\alpha}_s W_s'} e^{\widehat{\tau}_s M_s'} \widehat{G}_s e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s} W_s y,$$
$$a_3 = y' e^{\widehat{\alpha}_s W_s'} e^{\widehat{\tau}_s M_s'} \widehat{G}_s e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s} W_s^2 y, \quad A_1 = e^{\widehat{\alpha}_s W_s'} e^{\widehat{\tau}_s M_s'} \widehat{G}_s e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s} y,$$
$$A_2 = e^{\widehat{\alpha}_s W_s'} e^{\widehat{\tau}_s M_s'} \widehat{G}_s e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s} W_s y, \quad A_3 = W_s' e^{\widehat{\alpha}_s W_s'} e^{\widehat{\tau}_s M_s'} \widehat{G}_s e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s} y,$$

[5] Note that $\widetilde{P}_s$ is derived for a given $(\widehat{\alpha}_s, \widehat{\tau}_s)$ value. Thus, we assume that this term is a non-random variable in applying Lemma A.1 in Appendix A.

$$b_1 = y' e^{\widehat{\alpha}_s W_s'} e^{\widehat{\tau}_s M_s'} \left(2\widehat{G}_s M_s + \frac{\partial \widehat{G}_s}{\partial \widehat{\tau}_s} \right) e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s} y,$$

$$b_2 = y' e^{\widehat{\alpha}_s W_s'} e^{\widehat{\tau}_s M_s'} \left(2\widehat{G}_s M_s + \frac{\partial \widehat{G}_s}{\partial \widehat{\tau}_s} \right) M_s e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s} y,$$

$$b_3 = y' e^{\widehat{\alpha}_s W_s'} e^{\widehat{\tau}_s M_s'} \left(2\frac{\partial \widehat{G}_s}{\partial \widehat{\tau}_s} M_s + \frac{\partial^2 \widehat{G}_s}{\partial \widehat{\tau}_s^2} \right) e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s} y,$$

$$B_1 = e^{\widehat{\alpha}_s W_s'} e^{\widehat{\tau}_s M_s'} \widehat{G}_s e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s} y, \quad B_2 = e^{\widehat{\alpha}_s W_s'} e^{\widehat{\tau}_s M_s'} \left(2\widehat{G}_s M_s + \frac{\partial \widehat{G}_s}{\partial \widehat{\tau}_s} \right) e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s} y.$$

Here, $\frac{\partial \widehat{G}_s}{\partial \widehat{\tau}_s} = -\frac{\partial \widetilde{P}_s}{\partial \widehat{\tau}_s}$ and $\frac{\partial^2 \widehat{G}_s}{\partial \widehat{\tau}_s^2} = -e^{-\widehat{\alpha}_s W_s} e^{-\widehat{\tau}_s M_s} \big(\widehat{P}_s (M_s + M_s')(\widehat{G}_s M_s - \frac{\partial \widehat{P}_s}{\partial \widehat{\tau}_s}) + (\frac{\partial \widehat{P}_s}{\partial \widehat{\tau}_s} - M_s \widehat{P}_s)(M_s + M_s')\widehat{G}_s \big) e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s}$. Then, given an estimator $\widehat{\Omega}$[6], our feasible selection criterion function is consequently given by

$$\widehat{C}_s = \left\| \widetilde{P}_s y - y \right\|^2 + 2 \left(\operatorname{tr}\left(\widetilde{P}_s \widehat{\Omega} \right) + \frac{\partial \widehat{\alpha}_s}{\partial y'} \widehat{\Omega} \frac{\partial \widetilde{P}_s}{\partial \widehat{\alpha}_s} y + \frac{\partial \widehat{\tau}_s}{\partial y'} \widehat{\Omega} \frac{\partial \widetilde{P}_s}{\partial \widehat{\tau}_s} y \right). \tag{3.8}$$

Using $\widehat{C}_s$, the selected model is defined in the following way.

$$\widehat{s} = \operatorname{argmin}_{s \in \{1,\ldots,S\}} \widehat{C}_s. \tag{3.9}$$

Remark 1. *Consider the following linear regression model:* $y_i = \mu_i + \epsilon_i$, *where* $\mu_i = x_i'\beta$ *for* $i = 1, \ldots, n$. *Let* $X = (x_1, \ldots, x_n)'$ *and consider our criterion function in* (3.3). *In the context of this model, (i)* $\tilde{P}_s$ *reduces to* P, *where* $P = X(X'X)^{-1}X'$, *(ii)* Ω *reduces to* $\sigma^2 I_n$ *and (iii)* $\frac{\partial \widehat{\alpha}_s}{\partial y'} \Omega \frac{\partial \widetilde{P}_s}{\partial \widehat{\alpha}_s} y + \frac{\partial \widehat{\tau}_s}{\partial y'} \Omega \frac{\partial \widetilde{P}_s}{\partial \widehat{\tau}_s} y = 0$. *Thus, our criterion function in* (3.3) *reduces to the Mallows* C_p *formula (Mallows, 1973) given by* $\|Py - y\|^2 + 2k\sigma^2$. *Let* $\widehat{\mu} = Py = X(X'X)^{-1}X'y$ *be an estimator of* $\mu = (\mu_1, \ldots, \mu_n)'$, *and* y_i^0 *be i.i.d.*(μ_i, σ^2) *for* $i = 1, \ldots, n$. *Define* $Err_i = \mathbb{E}(y_i^0 - \widehat{\mu}_i)^2$ *and* $Err_i^* = \mathbb{E}(y_i^0 - \mu_i)^2$. *Then,*

$$\begin{aligned} Err_i &= \mathbb{E}(y_i^0 - \widehat{\mu}_i + \mu_i - \mu_i)^2 = \mathbb{E}(y_i^0 - \mu_i)^2 + \mathbb{E}(\widehat{\mu}_i - \mu_i)^2 - 2\mathbb{E}(\widehat{\mu}_i - \mu_i)(y_i^0 - \mu_i) \\ &= Err_i^* + R - 2\operatorname{Cov}(\widehat{\mu}_i, y_i), \end{aligned}$$

where $R = \mathbb{E}(\widehat{\mu}_i - \mu_i)^2$ *is the risk function. Hence,* $R = (Err_i - Err_i^*) + 2\operatorname{Cov}(\widehat{\mu}_i, y_i)$, *where the first term is the prediction accuracy. In order to unbiasedly estimate* R, *we need to account for the covariance term* $2\operatorname{Cov}(\widehat{\mu}_i, y_i)$ *as well, and the second term in* (3.3) *serves to this purpose (Mallows, 1973).*

Let (W^*, M^*) be the true spatial weights matrices. We will first assume that $\mathcal{W}$ does not

[6] Following Zhang and Yu (2018), we use the plug-in estimator of $\widehat{\Omega}$ given by $\widehat{\Omega} = \widehat{\sigma}_s^2 e^{-\widehat{\alpha}_s W_s} e^{-\widehat{\tau}_s M_s} e^{-\widehat{\tau}_s M_s'} e^{-\widehat{\alpha}_s W_s'}$ in our analysis, where $\widehat{\sigma}_s^2$, $\widehat{\alpha}_s$ and $\widehat{\tau}_s$ are the estimators of σ_s^2, α_s and τ_s, respectively.

include (W^*, M^*), and establish the asymptotic optimality of our selection procedure defined in (3.9). Assume that $\widehat{\alpha}_s$ and $\widehat{\tau}_s$ have the probability limits α_s^* and τ_s^*, respectively. Denote $P_s^* = \widetilde{P}_s|_{\widehat{\alpha}_s=\alpha_s^*, \widehat{\tau}_s=\tau_s^*}$, $\mu_s^* = P_s^* y$, $R_s^* = E\|\mu_s^* - \mu\|^2$ and $\zeta_n = \inf_{s\in\{1,\ldots,S\}} R_s^*$. Let $\gamma_{\min}(A)$ and $\gamma_{\max}(A)$ be the minimum and maximum eigenvalues of a matrix A, respectively. For notational convenience, let $\inf_s$ ($\sup_s$) denote the infimum (supremum) over $s \in \{1, \ldots, S\}$. We make the following assumptions.

Assumption 1. *The spatial weights matrices $\{W\}$ and $\{M\}$ are bounded in both row and column sum norms.*

Assumption 2. *The parameter space for α and τ is a compact subset of $\mathbb{R}^2$ and includes α^* and τ^*.*

Assumption 3. *(i) The matrix X has full column rank and its elements are uniformly bounded constants. (ii) For all values of τ in its compact parameter space, (a) $\lim_{n\to\infty} X' e^{\tau M'} e^{\tau M} X/n$ exists and is non-singular, and (b) $\gamma_{\min}(e^{\tau M'} e^{\tau M})$ is bounded away from zero uniformly.*

Assumption 4. *ϵ_i's are i.i.d. with mean zero and variance σ^2, and $\mathbb{E}(\epsilon_i^{4G})$ exists, where G is defined in Assumption 5.*

Assumption 5. *There exists a positive integer $G \geq 1$ such that $\sum_{s=1}^{S} (R_s^*)^{-G} = o(1)$.*

Assumption 6. $\|\mu\|^2 = O(n)$.

Assumption 7. $\zeta_n^{-1} \sup_s \left| \frac{\partial \widehat{\alpha}_s}{\partial y'} \widehat{\Omega} \frac{\partial \widetilde{P}_s}{\partial \widehat{\alpha}_s} y \right| = o_p(1)$ *and* $\zeta_n^{-1} \sup_s \left| \frac{\partial \widehat{\tau}_s}{\partial y'} \widehat{\Omega} \frac{\partial \widetilde{P}_s}{\partial \widehat{\tau}_s} y \right| = o_p(1)$.

Assumption 8. $\zeta_n^{-1} k = o(1)$ *and* $n\zeta_n^{-1} \sup_s \gamma_{\max}(\widetilde{P}_s - P_s^*) = o_p(1)$.

Assumption 1 is a standard assumption adopted in the spatial econometric literature (Kelejian and Prucha, 1998, 2010; Lee, 2003). It allows for limiting the spatial correlations between units so that the asymptotic analysis becomes manageable. Assumption 2 ensures that $e^{\alpha W}$ and $e^{\tau M}$ are bounded in both row and column sum norms. This can be seen from $\left\|e^{\alpha W}\right\| = \left\|\sum_{i=0}^{\infty} \frac{\alpha^i W^i}{i!}\right\| \leq \sum_{i=0}^{\infty} \frac{|\alpha|^i \|W\|^i}{i!} = e^{|\alpha| \|W\|}$ is bounded if $|\alpha|$ is bounded, where $\|\cdot\|$ is either the maximum row sum norm or the maximum column sum norm. Similarly, $\left\|e^{\tau W}\right\|$ is bounded if $|\tau|$ is bounded. Assumption 3 is a standard assumption adopted for the MESS, see e.g., Debarsy et al. (2015). Assumption 4 is also a standard assumption adopted in the literature on spatial econometrics, see e.g., Kelejian and Prucha (1998, 2010) and Lee (2003, 2004). It is used to show that some quadratic forms of the error terms required in the analysis are bounded in probability. This assumption is also common in the literature on the model selection optimality (Li, 1987; Zhang and Yu, 2018). We use this moment condition to bound linear and quadratic terms of $\tilde{\epsilon}$ in probability by Whittle's inequality (Whittle, 1960). Assumption 6 requires that $\mu' \mu/n = O(1)$, which is also a common condition assumed in the model selection and model averaging literature (Liang et al., 2011; Zhang and Yu, 2018).

To understand Assumption 7, we can use (3.4) and (3.6) to express $\frac{\partial \widehat{\alpha}_s}{\partial y'}\widehat{\Omega}\frac{\partial \widetilde{P}_s}{\partial \widehat{\alpha}_s}y$ as

$$\frac{\partial \widehat{\alpha}_s}{\partial y'}\widehat{\Omega}\frac{\partial \widetilde{P}_s}{\partial \widehat{\alpha}_s}y = -\frac{\left(2a_1A_1 - \left\|\widehat{G}_s e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s} y\right\|^2 (A_2 + A_3)\right)' \widehat{\Omega}\left(-W_s\widetilde{P}_s y + \widetilde{P}_s W_s y\right)}{2a_1^2 - \left\|\widehat{G}_s e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s} y\right\|^2 (a_2 + a_3)}. \tag{3.10}$$

Our Assumptions 1 and 2 ensure that $\widehat{\Omega} = O_p(1)$. All other terms in (3.10) are either linear or quadratic terms in y. These terms are $O_p(n)$ under some regularity conditions, e.g., see Debarsy et al. (2015). Hence, it follows that $\frac{\partial \widehat{\alpha}_s}{\partial y'}\widehat{\Omega}\frac{\partial \widetilde{P}_s}{\partial \widehat{\alpha}_s}y = O_p(n)$. Thus, Assumption 7 will be satisfied if $\zeta_n^{-1} = o(1/n)$. The second part of Assumption 7 can be justified similarly. The first part of Assumption 8 is a common assumption in the literature on model averaging (Liu and Okui, 2013; Zhang and Yu, 2018). So long as n increases faster than k, our Assumptions 1–3 ensures that $\zeta_n^{-1}k = o(1)$ in our setting.[7] The second part of Assumption 8 shows that the term $\zeta_n^{-1}\sup_s \gamma_{max}(\widetilde{P}_s - P_s^*)$ converges to zero in probability at a rate faster than n. This is also a common assumption in the model averaging literature (Zhang et al., 2014). Under our Assumptions 1–3, it can be shown that $\left\|\widetilde{P}_s - P_s^*\right\| = o_p(1)$, where $\|\cdot\|$ can be either the maximum row sum norm or the maximum column sum norm, which implies that $\sup_s \gamma_{max}(\widetilde{P}_s - P_s^*) = o_p(1)$.[8] Hence, Assumption 8 requires that $\zeta_n^{-1} = o(1/n)$, which is consistent with the order required by Assumption 7.

Theorem 1. *Assume that Assumptions 1–8 hold. Then, for any $\eta > 0$, it follows that*

$$\lim_{n\to\infty} P\left(\left|\frac{L_{\widehat{s}}}{\inf_{s\in\{1,\ldots,S\}} L_s} - 1\right| > \eta\right) = 0. \tag{3.11}$$

The proof of Theorem 1 is given in Appendix E.1. Theorem 1 indicates that the squared loss function of our selected $\widehat{\mu}_{\widehat{s}}$ estimator is asymptotically identical to the squared loss function of the infeasible estimator that uses the best candidate spatial weight matrix. That is, the selection estimator $\widehat{\mu}_{\widehat{s}}$ is asymptotically optimal in the sense that it is as efficient as the infeasible estimator that uses the best candidate spatial weights matrix.

Next, we assume that $\mathcal{W}$ includes (W^*, M^*) and show that our selection procedure defined in (3.9) is selection consistent. Let $\mathcal{S}$ be the index set of spatial weight matrices excluding the true tuple (W^*, M^*), i.e., $\mathcal{S} = \{s \in \{1,\ldots,S\} : (W_s, M_s) \neq (W^*, M^*)\}$. Define $\zeta_n^* = \inf_{s\in\mathcal{S}} R_s^*$. Then, our selection consistency result requires the following assumption.

[7] Note that we can write R_s^* as $R_s^* = \mathbb{E}\left\|\mu_s^* - \mu\right\|^2 = \mathrm{tr}(P_s^*\Omega P_s^{*'}) + \|H_s^*\mu\|^2$, where $H_s^* = I_n - P_s^*$. Using some arguments given in the proof of Theorem 1, it can be shown that $R_s^* = O(n+k)$. Thus, $\zeta_n^{-1}k = O(k/(n+k)) = o(1)$ if $n/k \to \infty$.

[8] Note that we can write $e^{\widehat{\alpha}_s W} = \left(e^{\widehat{\alpha}_s W} - e^{\alpha_s^* W}\right) + e^{\alpha_s^* W}$ and $e^{\widehat{\tau}_s M} = \left(e^{\widehat{\tau}_s M} - e^{\tau_s^* M}\right) + e^{\tau_s^* M}$. Also $\left\|e^{\widehat{\alpha}_s W} - e^{\alpha_s^* W}\right\|_1 = \left\|\left(e^{(\widehat{\alpha}_s - \alpha_s^*)W} - I_n\right)e^{\alpha_s^* W}\right\|_1 \leq \left\|\left(e^{(\widehat{\alpha}_s - \alpha_s^*)W} - I_n\right)\right\|_1 \left\|e^{\alpha_s^* W}\right\|_1 = o_p(1)$ by Lemmas A.3 and A.4 in Appendix A, where $\|\cdot\|_1$ is the maximum column sum matrix norm. Similarly, $\left\|e^{\widehat{\tau}_s M} - e^{\tau_s^* M}\right\|_1 = o_p(1)$. These observations can be used to show that $\left\|\widetilde{P}_s - P_s^*\right\|_1 = o_p(1)$.

Assumption 9. *If $\mathcal{S}$ is not empty, then we have (i)* $\sup_s \left| \frac{\partial \widehat{\alpha}_s}{\partial y'} \widehat{\Omega} \frac{\partial \widetilde{P}_s}{\partial \widehat{\alpha}_s} y \right| / \zeta_n^* = o_p(1)$ *and* $\sup_s \left| \frac{\partial \widehat{\tau}_s}{\partial y'} \widehat{\Omega} \frac{\partial \widetilde{P}_s}{\partial \widehat{\tau}_s} y \right| / \zeta_n^* = o_p(1)$*, (ii)* $n \sup_s \gamma_{\max}(\widetilde{P}_s - P_s^*)/\zeta_n^* = o_p(1)$*, and (iii)* $kn^{-1/2} = o(1)$ *and* $\zeta_n^{*-1} n = o(1)$*.*

The first and second part in Assumption 9 provide the counterparts of those conditions stated in Assumptions 7 and 8. The third part requires that (i) k may increase at a rate slower than $n^{1/2}$, and (ii) ζ_n^* may increase at a rate faster than n. These rates play an important role in bounding certain terms in our proof of selection consistency result.

Theorem 2. *Assume that Assumptions 1–4, 6 and 9 hold. Then,* $P\left((W_{\widehat{s}}, M_{\widehat{s}}) = (W^*, M^*)\right) \to 1$ *as* $n \to \infty$*.*

The proof of Theorem 2 is provided in Appendix E.2. Theorem 2 indicates that our selection procedure in (3.9) is selection consistent in the sense that it chooses the true tuple of weights matrices with probability approaching to one in large samples.

4 Model averaging procedure

Instead of selecting the asymptotically optimal model, we can use a model averaging scheme that compromises across a set of candidate models for the MESS(1,1). Compared with the model selection procedure in Section 3, the model averaging method can provide insurance against selecting an inappropriate model, and can reduce the risk associated with the loss function (Hansen, 2014). Denote the vector of model weights by $w = (w_1, \ldots, w_S)'$, and the set of model weights vectors by $\mathcal{N} = \left\{ w \in [0,1]^S : \sum_{s=1}^S w_s = 1 \right\}$. Let $\widetilde{P}(w) = \sum_{s=1}^S w_s \widetilde{P}_s$ be the weighted average of $\left\{ \widetilde{P}_1, \ldots, \widetilde{P}_S \right\}$. Then, the model average estimator for μ is given by

$$\widehat{\mu}(w) = \sum_{s=1}^S w_s \widehat{\mu}_s = \sum_{s=1}^S w_s \widetilde{P}_s y = \widetilde{P}(w) y. \tag{4.1}$$

The associated squared loss and its expectation are $L(w) = \|\widehat{\mu}(w) - \mu\|^2$ and $R(w) = \mathbb{E}\|\widehat{\mu}(w) - \mu\|^2$, respectively. Then, we consider the following model weights choice criterion function,

$$C(w) = \left\| \widetilde{P}(w) y - y \right\|^2 + 2 \left(\operatorname{tr}\left(\widetilde{P}(w) \Omega \right) + \sum_{s=1}^S w_s \left(\frac{\partial \widehat{\alpha}_s}{\partial y'} \Omega \frac{\partial \widetilde{P}_s}{\partial \widehat{\alpha}_s} y + \frac{\partial \widehat{\tau}_s}{\partial y'} \Omega \frac{\partial \widetilde{P}_s}{\partial \widehat{\tau}_s} y \right) \right). \tag{4.2}$$

Similar to C_s in (3.3), we can show that $\mathbb{E}(C(w)) = R(w) + \operatorname{tr}(\Omega)$ by using Lemma A.1 in Appendix A. The first term in (4.2) is a measure of goodness of fit. The second term is an unbiased estimator of $2\mathrm{Cov}(\widehat{\mu}(w), y)$, which can be called as the degrees of freedom of model averaging. Given an estimator $\widehat{\Omega}$, our feasible model weights choice criterion function can be expressed as

$$\widehat{C}(w) = \left\| \widetilde{P}(w) y - y \right\|^2 + 2 \left(\operatorname{tr}\left(\widetilde{P}(w) \widehat{\Omega} \right) + \sum_{s=1}^S w_s \left(\frac{\partial \widehat{\alpha}_s}{\partial y'} \widehat{\Omega} \frac{\partial \widetilde{P}_s}{\partial \widehat{\alpha}_s} y + \frac{\partial \widehat{\tau}_s}{\partial y'} \widehat{\Omega} \frac{\partial \widetilde{P}_s}{\partial \widehat{\tau}_s} y \right) \right). \tag{4.3}$$

The optimal model weights vector is thus given by

$$\widehat{w} = \text{argmin}_{w\in\mathcal{N}}\, \widehat{C}(w). \tag{4.4}$$

We can express $\widehat{C}(w)$ in a convenient form. Define $Q = \left(\widetilde{P}_1 y - y, \ldots, \widetilde{P}_S y - y\right)$ and

$$q = \begin{pmatrix} \text{tr}(\widetilde{P}_1\widehat{\Omega}) + \frac{\partial\widehat{\alpha}_1}{\partial y'}\widehat{\Omega}\frac{\partial\widetilde{P}_1}{\partial\widehat{\alpha}_1}y + \frac{\partial\widehat{\tau}_1}{\partial y'}\widehat{\Omega}\frac{\partial\widetilde{P}_1}{\partial\widehat{\tau}_1}y \\ \text{tr}(\widetilde{P}_2\widehat{\Omega}) + \frac{\partial\widehat{\alpha}_2}{\partial y'}\widehat{\Omega}\frac{\partial\widetilde{P}_2}{\partial\widehat{\alpha}_2}y + \frac{\partial\widehat{\tau}_2}{\partial y'}\widehat{\Omega}\frac{\partial\widetilde{P}_2}{\partial\widehat{\tau}_2}y \\ \vdots \\ \text{tr}(\widetilde{P}_S\widehat{\Omega}) + \frac{\partial\widehat{\alpha}_S}{\partial y'}\widehat{\Omega}\frac{\partial\widetilde{P}_S}{\partial\widehat{\alpha}_S}y + \frac{\partial\widehat{\tau}_S}{\partial y'}\widehat{\Omega}\frac{\partial\widetilde{P}_S}{\partial\widehat{\tau}_S}y \end{pmatrix}. \tag{4.5}$$

Then, we can express (4.3) as $\widehat{C}(w) = w'Q'Qw + 2w'q$, which indicates that the optimization problem in (4.4) is a quadratic programming (or a quadratic optimization) problem. Optimization of a quadratic form subject to some constraints is a well-studied area of numerical optimization, and numerical solvers are available in popular statistical software such as Python, R and Matlab.

Next, we will show that our MA estimator that uses (4.4) is asymptotically optimal in the sense that its associated squared loss is asymptotically equivalent to the smallest squared loss. Let $R^*(w) = \mathbb{E}\left\|\sum_{s=1} w_s\widehat{\mu}_s|_{\widehat{\alpha}_s=\alpha_s^*,\widehat{\tau}_s=\tau_s^*} - \mu\right\|^2$ and $\tilde{\zeta}_n = \inf_{w\in\mathcal{N}} R^*(w)$. To show the asymptotic optimality of the averaging estimator $\widehat{\mu}(\widehat{w})$, we adopt the following assumption.

Assumption 10. *(i) There exists a positive integer G such that $S\tilde{\zeta}_n^{-2G}\sum_{s=1}^S (R_s^*)^G = o(1)$, (ii) $\tilde{\zeta}_n^{-1}\sup_s\left|\frac{\partial\widehat{\alpha}_s}{\partial y'}\widehat{\Omega}\frac{\partial\widetilde{P}_s}{\partial\widehat{\alpha}_s}y\right| = o_p(1)$ and $\tilde{\zeta}_n^{-1}\sup_s\left|\frac{\partial\widehat{\tau}_s}{\partial y'}\widehat{\Omega}\frac{\partial\widetilde{P}_s}{\partial\widehat{\tau}_s}y\right| = o_p(1)$, (iii) $\tilde{\zeta}_n^{-1}k = o(1)$ and $n\tilde{\zeta}_n^{-1}\sup_s\gamma_{\max}(\widetilde{P}_s - P_s^*) = o_p(1)$.*

The first part of Assumption 10 is a key assumption and is common in the model averaging literature (Liu and Okui, 2013; Wan et al., 2010). Similar to our analysis in model selection, this part allows us to determine the probability orders of terms that are linear and quadratic in $\tilde{\epsilon}$ through the generalized Chebyshev's inequality and Whittle's inequality. It implicitly assumes a trade-off between the number of models allowed and the fit of individual models. The rest of Assumption 10 provides conditions that are counterparts to those adopted in Assumptions 7 and 8.

Theorem 3. *Assume that Assumptions 1–4, 6 and 10 hold. Then, for any $\eta > 0$, it follows that*

$$\lim_{n\to\infty} P\left(\left|\frac{L(\widehat{w})}{\inf_{w\in\mathcal{N}} L(w)} - 1\right| > \eta\right) = 0. \tag{4.6}$$

The proof of Theorem 3 is provided in Appendix E.3. Similar to the model selection procedure presented in Section 3, Theorem 3 indicates that the model averaging estimator is also asymptotically optimal. Since $\inf_{w\in\mathcal{N}} L(w) \leq \inf_s L_s$, the model averaging procedure can be superior over the model selection procedure in the sense that it could possibly reduce the squared loss.

5 Extensions

Before we consider some extensions of our model specification, it is easy to see that our results are also valid for two special cases of MESS$(1,1)$, namely, MESS$(1,0)$ and MESS$(0,1)$. In Appendix C, we provide the required quantities for the model selection and averaging procedures for these special cases. Next, we consider two important extensions of our model, and show how our suggested criterion functions can be formulated for these extensions. In the first extension, we consider a higher-order MESS model and derive the associated selection and weight choice criterion functions. In the second extension, we show how our analysis should be modified when the disturbance terms are heteroskedastic. We also show how to derive the selection and weight choice criterion functions in a heteroskedasticity-robust GMM framework.

5.1 The higher-order MESS

A high-order MESS model, namely MESS(p,q), can be formulated in the following way

$$e^{\sum_{i=1}^{p}\alpha_i W_i}y = X\beta + u, \quad e^{\sum_{j=1}^{q}\tau_j M_j}u = \epsilon, \tag{5.1}$$

where $\{W_i\}_{i=1}^{p}$ and $\{M_i\}_{i=1}^{q}$ are the spatial weights matrices, and $\{\alpha_i\}_{i=1}^{p}$ and $\{\tau_i\}_{i=1}^{q}$ are the unknown scalar spatial parameters. Here, we have $\mu = \mathbb{E}(y) = e^{-\sum_{i=1}^{p}\alpha_i W_i}X\beta$ and

$$\Omega = \sigma^2 e^{-\sum_{i=1}^{p}\alpha_i W_i}e^{-\sum_{j=1}^{q}\tau_j M_j}e^{-\sum_{j=1}^{q}\tau_j M_j'}e^{-\sum_{i=1}^{p}\alpha_i W_i'}.$$

Let W_s^i and M_s^j be the ith and jth spatial weights matrices of the MESS(p,q) in the sth model with corresponding coefficients α_s^i and τ_s^j. Then, the quasi log-likelihood function for the model s can be expressed as

$$\ell_s = -\frac{n}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}\left\| e^{\sum_{j=1}^{q}\tau_j M_s^j}(e^{\sum_{i=1}^{p}\alpha_i W_s^i}y - X\beta)\right\|^2. \tag{5.2}$$

The estimator for μ_s is

$$\begin{aligned}\widehat{\mu}_s &= e^{-\sum_{i=1}^{p}\widehat{\alpha}_s^i W_s^i}X\left(X' e^{\sum_{j=1}^{q}\widehat{\tau}_s^j M_s^{j'}}e^{\sum_{j=1}^{q}\widehat{\tau}_s^j M_s^j}X\right)^{-1}X' e^{\sum_{j=1}^{q}\widehat{\tau}_s^j M_s^{j'}}e^{\sum_{j=1}^{q}\widehat{\tau}_s^j M_s^j}e^{\sum_{i=1}^{p}\widehat{\alpha}_s^i W_s^i}y \\ &= e^{-\sum_{i=1}^{p}\widehat{\alpha}_s^i W_s^i}e^{-\sum_{j=1}^{q}\widehat{\tau}_s^j M_s^j}\widehat{P}_s e^{\sum_{j=1}^{q}\widehat{\tau}_s^j M_s^j}e^{\sum_{i=1}^{p}\widehat{\alpha}_s^i W_s^i}y = \widetilde{P}_s y,\end{aligned} \tag{5.3}$$

where $\widehat{P}_s = e^{\sum_{j=1}^{q}\widehat{\tau}_s^j M_s^j}X\left(X' e^{\sum_{j=1}^{q}\widehat{\tau}_s^j M_s^{j'}}e^{\sum_{j=1}^{q}\widehat{\tau}_s^j M_s^j}X\right)^{-1}X' e^{\sum_{j=1}^{q}\widehat{\tau}_s^j M_s^{j'}}$. The feasible selection criterion function then takes the following form,

$$\widehat{C}_s = \left\|\widetilde{P}_s y - y\right\|^2 + 2\left(\operatorname{tr}\left(\widetilde{P}_s\widehat{\Omega}\right) + \sum_{i=1}^{p}\frac{\partial\widehat{\alpha}_s^i}{\partial y'}\widehat{\Omega}\frac{\partial\widetilde{P}_s}{\partial\widehat{\alpha}_s^i}y + \sum_{j=1}^{q}\frac{\partial\widehat{\tau}_s^j}{\partial y'}\widehat{\Omega}\frac{\partial\widetilde{P}_s}{\partial\widehat{\tau}_s^j}y\right). \tag{5.4}$$

Since ℓ_s in (5.2) is free of the Jacobian terms, the closed forms of the quantities required for the computation of (5.4) can be obtained conveniently. These expressions are

$$\frac{\partial \widetilde{P}_s}{\partial \widehat{\alpha}_s^i} = -W_s^i \widetilde{P}_s + \widetilde{P}_s W_s^i, \tag{5.5}$$

$$\frac{\partial \widetilde{P}_s}{\partial \widehat{\tau}_s^j} = e^{-\sum_{i=1}^p \widehat{\alpha}_s^i W_s^i} e^{-\sum_{j=1}^q \widehat{\tau}_s^j M_s^j} \widehat{P}_s \left(M_s^j + M_s^{j'}\right) \left(I_n - \widehat{P}_s\right) e^{\sum_{j=1}^q \widehat{\tau}_s^j M_s^j} e^{\sum_{i=1}^p \widehat{\alpha}_s^i W_s^i}, \tag{5.6}$$

and

$$\frac{\partial \widehat{\alpha}_s^i}{\partial y} = -\frac{2c_1 C_1 - \left\| \widehat{G}_s e^{\sum_{j=1}^q \widehat{\tau}_s^j M_s^j} e^{\sum_{i=1}^p \widehat{\alpha}_s^i W_s^i} y \right\|^2 (C_2 + C_3)}{2c_1^2 - \left\| \widehat{G}_s e^{\sum_{j=1}^q \widehat{\tau}_s^j M_s^j} e^{\sum_{i=1}^p \widehat{\alpha}_s^i W_s^i} y \right\|^2 (c_2 + c_3)}, \tag{5.7}$$

$$\frac{\partial \widehat{\tau}_s^j}{\partial y} = -\frac{2d_1 D_1 - 2\left\| \widehat{G}_s e^{\sum_{j=1}^q \widehat{\tau}_s^j M_s^j} e^{\sum_{i=1}^p \widehat{\alpha}_s^i W_s^i} y \right\|^2 D_2}{d_1^2 - \left\| \widehat{G}_s e^{\sum_{j=1}^q \widehat{\tau}_s^j M_s^j} e^{\sum_{i=1}^p \widehat{\alpha}_s^i W_s^i} y \right\|^2 (2d_2 + d_3)}, \tag{5.8}$$

where

$$c_1 = y' e^{\sum_{i=1}^p \widehat{\alpha}_s^i W_s^{i'}} e^{\sum_{j=1}^q \widehat{\tau}_s^j M_s^{j'}} \widehat{G}_s e^{\sum_{j=1}^q \widehat{\tau}_s^j M_s^j} e^{\sum_{i=1}^p \widehat{\alpha}_s^i W_s^i} W_s^i y,$$
$$c_2 = y' W_s^{i'} e^{\sum_{i=1}^p \widehat{\alpha}_s^i W_s^{i'}} e^{\sum_{j=1}^q \widehat{\tau}_s^j M_s^{j'}} \widehat{G}_s e^{\sum_{j=1}^q \widehat{\tau}_s^j M_s^j} e^{\sum_{i=1}^p \widehat{\alpha}_s^i W_s^i} W_s^i y,$$
$$c_3 = y' e^{\sum_{i=1}^p \widehat{\alpha}_s^i W_s^{i'}} e^{\sum_{j=1}^q \widehat{\tau}_s^j M_s^{j'}} \widehat{G}_s e^{\sum_{j=1}^q \widehat{\tau}_s^j M_s^j} e^{\sum_{i=1}^p \widehat{\alpha}_s^i W_s^i} (W_s^i)^2 y,$$

$$C_1 = e^{\sum_{i=1}^p \widehat{\alpha}_s^i W_s^{i'}} e^{\sum_{j=1}^q \widehat{\tau}_s^j M_s^{j'}} \widehat{G}_s e^{\sum_{j=1}^q \widehat{\tau}_s^j M_s^j} e^{\sum_{i=1}^p \widehat{\alpha}_s^i W_s^i} y,$$
$$C_2 = e^{\sum_{i=1}^p \widehat{\alpha}_s^i W_s^{i'}} e^{\sum_{j=1}^q \widehat{\tau}_s^j M_s^{j'}} \widehat{G}_s e^{\sum_{j=1}^q \widehat{\tau}_s^j M_s^j} e^{\sum_{i=1}^p \widehat{\alpha}_s^i W_s^i} W_s^i y,$$
$$C_3 = W_s^{i'} e^{\sum_{i=1}^p \widehat{\alpha}_s^i W_s^{i'}} e^{\sum_{j=1}^q \widehat{\tau}_s^j M_s^{j'}} \widehat{G}_s e^{\sum_{j=1}^q \widehat{\tau}_s^j M_s^j} e^{\sum_{i=1}^p \widehat{\alpha}_s^i W_s^i} y,$$

$$d_1 = y' e^{\sum_{i=1}^p \widehat{\alpha}_s^i W_s^{i'}} e^{\sum_{j=1}^q \widehat{\tau}_s^j M_s^{j'}} \left(2\widehat{G}_s M_s^j + \frac{\partial \widehat{G}_s}{\partial \widehat{\tau}_s^j}\right) e^{\sum_{j=1}^q \widehat{\tau}_s^j M_s^j} e^{\sum_{i=1}^p \widehat{\alpha}_s^i W_s^i} y,$$
$$d_2 = y' e^{\sum_{i=1}^p \widehat{\alpha}_s^i W_s^{i'}} e^{\sum_{j=1}^q \widehat{\tau}_s^j M_s^{j'}} \left(2\widehat{G}_s M_s^j + \frac{\partial \widehat{G}_s}{\partial \widehat{\tau}_s^j}\right) M_s^j e^{\sum_{j=1}^q \widehat{\tau}_s^j M_s^j} e^{\sum_{i=1}^p \widehat{\alpha}_s^i W_s^i} y,$$
$$d_3 = y' e^{\sum_{i=1}^p \widehat{\alpha}_s^i W_s^{i'}} e^{\sum_{j=1}^q \widehat{\tau}_s^j M_s^{j'}} \left(2\frac{\partial \widehat{G}_s}{\partial \widehat{\tau}_s^j} M_s^j + \frac{\partial^2 \widehat{G}_s}{\partial \widehat{\tau}_s^{j2}}\right) e^{\sum_{j=1}^q \widehat{\tau}_s^j M_s^j} e^{\sum_{i=1}^p \widehat{\alpha}_s^i W_s^i} y,$$

$$D_1 = e^{\sum_{i=1}^{p}\widehat{\alpha}_s^i W_s^{i'}} e^{\sum_{j=1}^{q}\widehat{\tau}_s^j M_s^{j'}} \widehat{G}_s e^{\sum_{j=1}^{q}\widehat{\tau}_s^j M_s^j} e^{\sum_{i=1}^{p}\widehat{\alpha}_s^i W_s^i} y \quad \text{and}$$

$$D_2 = e^{\sum_{i=1}^{p}\widehat{\alpha}_s^i W_s^{i'}} e^{\sum_{j=1}^{q}\widehat{\tau}_s^j M_s^{j'}} \left(2\widehat{G}_s M_s^j + \frac{\partial \widehat{G}_s}{\partial \widehat{\tau}_s^j}\right) e^{\sum_{j=1}^{q}\widehat{\tau}_s^j M_s^j} e^{\sum_{i=1}^{p}\widehat{\alpha}_s^i W_s^i} y,$$

for $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, q$. Here, $\frac{\partial \widehat{G}_s}{\partial \widehat{\tau}_s^j} = -\frac{\partial \widetilde{P}_s}{\partial \widehat{\tau}_s^j}$ is given by the negative of (5.6) and $\frac{\partial^2 \widehat{G}_s}{\partial \widehat{\tau}_s^{j2}} = e^{-\sum_{i=1}^{p}\widehat{\alpha}_s^i W_s^i} e^{-\sum_{j=1}^{q}\widehat{\tau}_s^j M_s^j} \big(\widehat{P}_s(M_s^j + M_s^{j'})\big((I_n - \widehat{P}_s)M_s^j - \frac{\partial \widehat{P}_s}{\partial \widehat{\tau}_s^j}\big) + (\frac{\partial \widehat{P}_s}{\partial \widehat{\tau}_s^j} - M_s^j \widehat{P}_s)(M_s^j + M_s^{j'})(I_n - \widehat{P}_s)\big) e^{\sum_{j=1}^{q}\widehat{\tau}_s^j M_s^j} e^{\sum_{i=1}^{p}\widehat{\alpha}_s^i W_s^i}$.

Next, we consider the model averaging procedure. The model averaging estimator is given by $\widehat{\mu}(w) = \widetilde{P}(w)y = \sum_{s=1}^{S} w_s \widetilde{P}_s y$, and the feasible weights choice criterion function takes the following form,

$$\widehat{C}(w) = \left\|\widetilde{P}(w)y - y\right\|^2 + 2\left(\operatorname{tr}\left(\widetilde{P}(w)\widehat{\Omega}\right) + \sum_{s=1}^{S} w_s \left(\sum_{i=1}^{p} \frac{\partial \widehat{\alpha}_s^i}{\partial y'}\widehat{\Omega}\frac{\partial \widetilde{P}_s}{\partial \widehat{\alpha}_s^i} y + \sum_{j=1}^{q} \frac{\partial \widehat{\tau}_s^j}{\partial y'}\widehat{\Omega}\frac{\partial \widetilde{P}_s}{\partial \widehat{\tau}_s^j} y\right)\right). \quad (5.9)$$

We conjecture that our results in Theorems 1–3 can be extended to the MESS(p, q) model under some regularity assumptions extended from those we adopted for the MESS$(1, 1)$ model.

5.2 The MESS with an unknown form of heteroskedasticity

In this section, we consider the MESS(1,1) with heteroskedastic error terms, i.e., $\epsilon_i \sim (0, \sigma_i^2)$. Let $\Sigma = \operatorname{Diag}(\sigma_1^2, \ldots, \sigma_n^2)$ be the diagonal $n \times n$ matrix. The quasi MLE of the MESS(1,1) can be consistent under heteroskedasticity when the spatial weights matrices W and M are commutative, i.e., $WM = MW$ (Debarsy et al., 2015).[9] Thus, we can use the quasi ML approach discussed in Sections 3 and 4 for the MS and MA procedures. Under heteroskedasticity, the variance of y becomes $\Omega = e^{-\alpha W} e^{-\tau M} \Sigma e^{-\tau M'} e^{-\alpha W'}$. The rest of the discussions in Sections 3 and 4 follow similarly. Thus, our results in Theorems 1–3 can be extended to the MESS(1,1) with the heteroskedastic error terms.

It is also possible to consider the GMM approach to extend our results in Theorems 1–3 to the heteroskedastic case. A consistent GMM estimator (GMME) can be based on a set of linear and quadratic moment functions. Let F be an $n \times k_f$ instrumental variable matrix, and P_i be an $n \times n$ matrix with zero diagonal elements for $i = 1, \ldots, k_p$. Then, the linear moment conditions are given by $F'\epsilon(\gamma)$, and the quadratic moments functions are given by $\epsilon'(\gamma)P_i\epsilon(\gamma)$ for $i = 1, \ldots, k_p$. Thus, we can formulate a GMME based on the following vector of the linear and quadratic moment functions[10]

$$g(\gamma) = \frac{1}{n}\left(\epsilon'(\gamma)P_1\epsilon(\gamma), \ldots, \epsilon'(\gamma)P_{k_p}\epsilon(\gamma), \epsilon'(\gamma)F\right)', \quad (5.10)$$

[9] In practice, it is often assumed that the spatial weights matrices for the dependent variable and the error terms are the same, i.e., $W = M$. In this case, commutativity is trivially satisfied.

[10] We do not focus on the selection of moment functions as our aim is the model selection and averaging procedures. Under certain conditions, Debarsy et al. (2015) show which F and P_i's can lead to the most efficient GMME.

where $\gamma = (\alpha, \tau, \beta')'$ and $\epsilon(\gamma) = e^{\tau M}(e^{\alpha W} y - X\beta)$. Let $\widetilde{\alpha}_s, \widetilde{\tau}_s$ and $\widetilde{\beta}_s$ be the GMMEs based on the sth model. Then, we can formulate the estimator of μ_s as $\widetilde{\mu}_s = e^{-\widetilde{\alpha}_s W_s} X \widetilde{\beta}_s$. Then, the feasible version of our selection criterion function is given by

$$\widetilde{C}_s = \|\widetilde{\mu}_s - y\|^2 + 2\text{tr}\left(\frac{\partial \widetilde{\mu}_s}{\partial y'}\widehat{\Omega}\right), \tag{5.11}$$

where $\frac{\partial \tilde{\mu}_s}{\partial y'} = \widetilde{P}_s + \frac{\partial \widetilde{P}_s}{\partial \tilde{\alpha}_s} y \frac{\partial \tilde{\alpha}_s}{\partial y'} + \frac{\partial \widetilde{P}_s}{\partial \tilde{\tau}_s} y \frac{\partial \tilde{\tau}_s}{\partial y'}$. The expressions for $\frac{\partial \tilde{\alpha}_s}{\partial y'}$ and $\frac{\partial \tilde{\tau}_s}{\partial y'}$ are stated in Appendix D. Define $\widetilde{s} = \text{argmin}_{s \in \{1,\ldots,S\}} \widetilde{C}_s$ as the selected model. Note that we can also define a model weights choice criterion similar to (5.11) for the MA procedure. Under some assumptions similar to our adopted ones, we conjecture that the results in Theorems 1–3 can be extended to the robust GMM case.

6 Monte Carlo simulations

In this section, we investigate the finite sample performance of the proposed MS and MA procedures. We consider the data generating process described below,

$$e^{\alpha W} y = \beta_1 x_1 + \beta_2 x_2 + u, \quad e^{\tau W} u = \epsilon. \tag{6.1}$$

The elements of x_1 and x_2 are independently drawn from $U(0, \sqrt{12})$ and $N(0, 1)$, respectively. The candidate spatial weights matrices W_1, W_2, W_3 and W_4 are selected as the true weights matrix in different cases. Here, W_1 is a square matrix with its elements interacting with only their left neighbors. For the left edge units, they interact with their right neighbors. W_2 is a square weights matrix with its elements interacting their left and right neighbors. W_3 and W_4 are based on rook and queen contiguity, respectively. To this end, n spatial units are randomly allocated into $\sqrt{n} \times \sqrt{n}$ square lattice graph. In the rook contiguity case, $w_{ij} = 1$ if the j'th observation is adjacent (left/right/above or below) to the i'th observation on the graph. In the queen contiguity case, $w_{ij} = 1$ if the j'th observation is adjacent to, or shares a border with the i'th observation. All weights matrices are row normalized. Note also that the spatial weights matrices become denser from W_1 to W_4. We set $(\beta_1, \beta_2)' = (2, 1)'$, and let α and τ take values from $\{-1.2, -0.2, 0.2, 1.2\}$. For the error terms we have three specifications: (i) $\epsilon_i \sim$ i.i.d. $N(0, 1)$, (ii) $\epsilon_i \sim$ i.i.d. χ^2_3, where χ^2_3 is the chi-squared distribution with 3 degrees of freedom, (iii) $\epsilon_i = \eta_i x_{1i}$ where $\eta_i \sim$ i.i.d. χ^2_3. These three specifications represent the normal, non-normal and heteroskedastic cases, respectively. In cases (ii) and (iii), ϵ_i is standardized so that its mean is zero and its variance is one. We consider two sample sizes, $n = 169$ and $n = 400$. We set the number of repetitions to 1000 in all cases.

We compute the QMLE and report the root mean squared error (RMSE) for each parameter under each spatial weights matrix and the selected weights matrix (MS). After implementing the model selection using our criterion function in (3.8), we compute the average of the loss function $L_s = \|\widehat{\mu}_s - \mu\|^2$ over 1000 repetitions under each spatial weights matrix. This measure is denoted by

"Loss" in the tables. The average loss function value for MS and MA are also reported. We compute the selection frequency under each spatial weights matrix, which is denoted by "MS accuracy". In each iteration, we also compute the model weight assigned to each model by our model averaging procedure. We then report the average of these model weights over 1000 repetitions. This measure is denoted by "MA weights" in the tables.

Table 1 shows the simulation results when the true weights matrix is W_1 and the error terms have the normal distribution. As expected, the RMSE is the smallest under W_1 for all parameters in all cases. The smallest loss is obtained by the estimator using W_1 in all cases. Specifically, when $\alpha = \tau = 0.2$ and $n = 169$, the value of the loss function under W_1 is 3.223, which is smaller than those reported from other spatial weights matrices (21.3 under W_2, 29.087 under W_3 and 33.987 under W_4). The loss for the MS estimator (3.501) and the MA estimator (3.823) are slightly larger than that of the QMLE using W_1. For the selection frequency, out of 1000 samples, the MS procedure selects W_1 985 times (98.5%), W_2 12 times (0.012%), W_3 3 times (0.003%) and W_4 0 times. For the MA weights, our model averaging procedure assigns an average weight of 88.4% to the true W_1, 5.9% to W_2, 3.5% to W_3 and 2.3% to W_4. Overall, these results imply that the MS procedure is picking the true spatial weights matrix, and the MA procedure is giving largest model weight to the one with the true spatial weights matrix. These results are consistent with our theoretical results.

When the spatial coefficients become negative, i.e., $\alpha = \tau = -0.2$, the results remain similar. The RMSE and loss are still smallest for W_1 and the MS accuracy and MA weights are also the largest under W_1. When the spatial parameters get larger in absolute value, i.e., $\alpha = \tau = 1.2$ or -1.2, the RMSE measures under W_1 are similar to other cases for α and τ, and smaller for β_1 and β_2, but become larger for all parameters under W_2, W_3 and W_4. The values of loss function become larger under all spatial weights matrices, but the one associated with W_1 remains the smallest. When $\alpha = \tau = 1.2$ and $n = 169$, the MS accuracy is 0.998 under W_1, and when $\alpha = \tau = -1.2$ and $n = 169$, it is 0.968 under W_1. When $\alpha = \tau = 1.2$ and $n = 169$, the MA weight assigned to W_1 is 0.915, and when $\alpha = \tau = -1.2$ and $n = 169$, the MA weight assigned to W_1 is 0.865. As n grows from 169 to 400, the MS and MA procedure provide more precise results in terms of selecting the true spatial weights matrix. In particular, when $\alpha = \tau = 0.2$ and $n = 400$, the MS selects W_1 1000 times (100%) and the other candidates 0 times. The MA weights grows to 92.7% for W_1, with 4%, 2.1% and 1.1% for W_2, W_3 and W_4, respectively.

Table 2 shows the simulation results when the true weights matrix is W_4 and the error terms have the normal distribution. The simulation results are similar to those reported in Table 1. The RMSE measures obtain the smallest values under W_4 (with an exception of τ when $n = 169$), and the smallest loss value occurs under W_4. The MS accuracy is always the largest under W_4, and the MA procedure assigns the largest weight to W_4 in all cases. When the true weights matrices are W_2 and W_3 and the error terms have the normal distribution, we observe similar simulation results to those reported Tables 1 and 2. For the sake of brevity, we provide those simulation results in the accompanying web appendix.

Table 3 reports the simulation results when the true wights matrix is W_1 and the error terms are non-normal, i.e., standardized χ^2_3. In general, the simulation results in this table are similar to those reported Table 1. The table reports the smallest RMSE and loss measures under W_1. Our MS procedure assigns the largest MS accuracy to W_1, and the MA procedure assigns the largest model weight to W_1 in all cases. When the true weights matrices are W_2, W_3 and W_4 and and the error terms are non-normal, we receive similar simulation results, which are provided in the accompanying web appendix. Table 4 reports the simulation results when the true weights matrix is W_1 and the error terms are heteroskedastic. The results in this table are similar to those reported for the normal and non-normal cases. These results indicate that our MS and MA procedure can be extended to the heteroskedastic errors case, as argued in Section 5.2. The results when W_2, W_3 and W_4 are true weights matrices and the error terms are heteroskedastic are similar, and are provided in the accompanying web appendix.

In Tables 1 to 4, we consider the cases where the set of candidate weights matrices includes the true spatial weights matrix. In Table 5, we consider a case where the true spatial weights matrix is not in the set of candidate weights matrices. To that end, we set the true spatial weights matrix to the sum of W_2 and W_4, which is not in the set of candidate weights matrices. The results in the table show that the losses are relatively small for W_2, W_3 and W_4 in most cases, except for the case in which $\alpha = \tau = 1.2$, where the loss for W_3 is bigger than W_1. This is reasonable because the true spatial weights matrix is given by $W_2 + W_4$. Since the set of candidate weights matrices does not include the true spatial weights matrix, the MA has smaller loss than MS in most cases, except when $\alpha = \tau = 1.2$ and $n = 400$. The MS procedure assigns relatively large MS accuracy to W_2 and W_4 as expected. Similarly, the MA procedure assigns relatively large weights to W_2 and W_4.

Table 1: Simulation results when W_1 is the true matrix and error terms are normal

			W_1	W_2	W_3	W_4	MS	MA
α=0.2	n=169	RMSE of α	0.034	0.046	0.063	0.082	0.034	
τ=0.2		RMSE of τ	0.087	0.094	0.137	0.201	0.086	
β_1=2		RMSE of β_1	0.074	0.079	0.078	0.082	0.074	
β_2=1		RMSE of β_2	0.079	0.079	0.083	0.084	0.079	
		Loss	3.223	21.300	29.087	33.987	3.501	3.823
		MS accuracy	0.985	0.012	0.003	0.000		
		MA weights	0.884	0.059	0.035	0.023		
α=0.2	n=400	RMSE of α	0.023	0.035	0.046	0.057	0.023	
τ=0.2		RMSE of τ	0.056	0.064	0.091	0.152	0.056	
β_1=2		RMSE of β_1	0.049	0.055	0.067	0.066	0.049	
β_2=1		RMSE of β_2	0.049	0.050	0.059	0.063	0.049	
		Loss	3.058	42.377	59.593	68.753	3.058	3.593
		MS accuracy	1.000	0.000	0.000	0.000		
		MA weights	0.927	0.040	0.021	0.011		
			W_1	W_2	W_3	W_4	MS	MA
α=−0.2	n=169	RMSE of α	0.033	0.046	0.075	0.124	0.038	
τ=−0.2		RMSE of τ	0.089	0.095	0.143	0.247	0.094	
β_1=2		RMSE of β_1	0.080	0.080	0.105	0.118	0.080	
β_2=1		RMSE of β_2	0.072	0.082	0.075	0.082	0.072	
		Loss	3.350	21.397	30.946	36.992	3.952	4.232
		MS accuracy	0.973	0.022	0.004	0.001		
		MA weights	0.870	0.070	0.026	0.035		
α=−0.2	n=400	RMSE of α	0.023	0.034	0.044	0.080	0.023	
τ=−0.2		RMSE of τ	0.056	0.063	0.091	0.191	0.056	
β_1=2		RMSE of β_1	0.050	0.057	0.055	0.055	0.050	
β_2=1		RMSE of β_2	0.049	0.050	0.052	0.056	0.049	
		Loss	3.156	42.508	61.742	73.335	3.214	3.755
		MS accuracy	0.999	0.001	0.000	0.000		
		MA weights	0.922	0.045	0.015	0.018		
			W_1	W_2	W_3	W_4	MS	MA
α=1.2	n=169	RMSE of α	0.031	0.472	7.283	6.186	0.036	
τ=1.2		RMSE of τ	0.085	0.401	7.182	6.294	0.087	
β_1=2		RMSE of β_1	0.064	0.848	1.221	0.898	0.075	
β_2=1		RMSE of β_2	0.048	0.423	0.624	0.503	0.052	
		Loss	8.637	1372.106	1784.061	1777.567	11.431	38.715
		MS accuracy	0.998	0.002	0.000	0.000		
		MA weights	0.915	0.032	0.026	0.027		
α=1.2	n=400	RMSE of α	0.021	0.430	4.906	3.148	0.021	
τ=1.2		RMSE of τ	0.054	0.283	4.923	3.172	0.054	
β_1=2		RMSE of β_1	0.043	0.825	0.843	0.612	0.043	
β_2=1		RMSE of β_2	0.035	0.444	0.447	0.333	0.035	
		Loss	9.722	3153.266	3706.969	3620.566	9.722	43.844
		MS accuracy	1.000	0.000	0.000	0.000		
		MA weights	0.937	0.028	0.020	0.014		
			W_1	W_2	W_3	W_4	MS	MA
α=−1.2	n=169	RMSE of α	0.030	0.359	6.061	12.592	0.617	
τ=−1.2		RMSE of τ	0.084	0.696	5.884	12.302	0.755	
β_1=2		RMSE of β_1	0.062	0.668	1.079	1.407	0.203	
β_2=1		RMSE of β_2	0.047	0.459	0.680	0.812	0.129	
		Loss	8.197	1032.113	1576.411	1886.296	57.850	76.213
		MS accuracy	0.968	0.022	0.002	0.008		
		MA weights	0.865	0.054	0.021	0.061		
α=−1.2	n=400	RMSE of α	0.018	0.420	6.802	12.104	0.018	
τ=−1.2		RMSE of τ	0.054	0.411	6.540	11.777	0.054	
β_1=2		RMSE of β_1	0.040	0.922	1.218	1.300	0.040	
β_2=1		RMSE of β_2	0.030	0.405	0.625	0.715	0.030	
		Loss	6.356	2942.293	3756.055	4063.768	6.356	49.286
		MS accuracy	1.000	0.000	0.000	0.000		
		MA weights	0.933	0.024	0.012	0.032		

Table 2: Simulation results when W_4 is the true matrix and error terms are normal

			W_1	W_2	W_3	W_4	MS	MA
α=0.2	n=169	RMSE of α	0.173	0.145	0.097	0.081	0.088	
τ=0.2		RMSE of τ	0.177	0.208	0.161	0.182	0.193	
β_1=2		RMSE of β_1	0.081	0.080	0.079	0.079	0.080	
β_2=1		RMSE of β_2	0.080	0.080	0.080	0.078	0.079	
		Loss	7.307	6.461	5.376	2.838	4.099	3.629
		MS accuracy	0.048	0.101	0.147	0.704		
		MA weights	0.085	0.121	0.140	0.653		
α=0.2	n=400	RMSE of α	0.177	0.153	0.091	0.056	0.063	
τ=0.2		RMSE of τ	0.158	0.184	0.133	0.115	0.130	
β_1=2		RMSE of β_1	0.051	11.774	0.051	0.050	0.050	
β_2=1		RMSE of β_2	0.051	0.051	0.050	0.050	0.050	
		Loss	11.774	10.497	7.706	2.738	4.073	3.689
		MS accuracy	0.018	0.054	0.108	0.820		
		MA weights	0.061	0.080	0.119	0.740		
			W_1	W_2	W_3	W_4	MS	MA
$\alpha=-0.2$	n=169	RMSE of α	0.179	0.157	0.118	0.095	0.111	
$\tau=-0.2$		RMSE of τ	0.121	0.143	0.187	0.186	0.186	
β_1=2		RMSE of β_1	0.079	0.079	0.078	0.078	0.078	
β_2=1		RMSE of β_2	0.073	0.074	0.074	0.072	0.074	
		Loss	7.827	7.261	5.485	3.425	4.835	4.329
		MS accuracy	0.069	0.083	0.286	0.562		
		MA weights	0.105	0.105	0.274	0.516		
$\alpha=-0.2$	n=400	RMSE of α	0.176	0.152	0.123	0.070	0.083	
$\tau=-0.2$		RMSE of τ	0.110	0.134	0.129	0.124	0.126	
β_1=2		RMSE of β_1	0.050	0.049	0.050	0.049	0.049	
β_2=1		RMSE of β_2	0.052	0.051	0.051	0.051	0.051	
		Loss	12.686	11.441	8.536	3.493	4.824	4.407
		MS accuracy	0.029	0.047	0.151	0.773		
		MA weights	0.077	0.077	0.159	0.687		
			W_1	W_2	W_3	W_4	MS	MA
α=1.2	n=169	RMSE of α	1.071	0.936	0.368	0.064	0.100	
τ=1.2		RMSE of τ	1.248	1.396	0.892	0.178	0.227	
β_1=2		RMSE of β_1	0.180	0.151	0.193	0.082	0.083	
β_2=1		RMSE of β_2	0.114	0.110	0.122	0.067	0.067	
		Loss	110.078	94.680	211.664	3.569	4.577	6.702
		MS accuracy	0.000	0.009	0.001	0.990		
		MA weights	0.027	0.058	0.029	0.886		
α=1.2	n=400	RMSE of α	1.043	0.911	0.315	0.043	0.043	
τ=1.2		RMSE of τ	1.293	1.434	0.842	0.112	0.112	
β_1=2		RMSE of β_1	0.148	0.125	0.204	0.054	0.054	
β_2=1		RMSE of β_2	0.082	0.076	0.083	0.046	0.046	
		Loss	250.292	209.017	486.556	3.775	3.775	6.916
		MS accuracy	0.000	0.000	0.000	1.000		
		MA weights	0.020	0.039	0.018	0.923		
			W_1	W_2	W_3	W_4	MS	MA
$\alpha=-1.2$	n=169	RMSE of α	1.094	0.805	1.230	0.144	0.402	
$\tau=-1.2$		RMSE of τ	0.241	0.516	0.318	0.219	0.265	
β_1=2		RMSE of β_1	0.280	0.213	0.325	0.090	0.112	
β_2=1		RMSE of β_2	0.150	0.147	0.179	0.071	0.080	
		Loss	325.396	220.216	364.073	18.729	57.196	45.227
		MS accuracy	0.032	0.110	0.017	0.841		
		MA weights	0.058	0.062	0.081	0.800		
$\alpha=-1.2$	n=400	RMSE of α	1.087	0.791	1.160	0.090	0.160	
$\tau=-1.2$		RMSE of τ	0.239	0.529	0.190	0.134	0.146	
β_1=2		RMSE of β_1	0.282	0.128	0.281	0.057	0.061	
β_2=1		RMSE of β_2	0.154	0.086	0.141	0.047	0.049	
		Loss	689.363	456.370	711.128	14.968	25.696	32.130
		MS accuracy	0.003	0.015	0.003	0.979		
		MA weights	0.025	0.026	0.051	0.898		

Table 3: Simulation results when W_1 is the true matrix and error terms are non-normal (χ_3^2)

			W_1	W_2	W_3	W_4	MS	MA
α=0.2	n=169	RMSE of α	0.032	0.047	0.061	0.084	0.034	
τ=0.2		RMSE of τ	0.087	0.095	0.132	0.187	0.087	
β_1=2		RMSE of β_1	0.076	0.080	0.079	0.084	0.075	
β_2=1		RMSE of β_2	0.080	0.080	0.084	0.087	0.080	
		Loss	3.221	21.377	29.073	34.087	3.665	3.818
		MS accuracy	0.977	0.020	0.002	0.001		
		MA weights	0.888	0.058	0.031	0.023		
α=0.2	n=400	RMSE of α	0.022	0.031	0.047	0.063	0.022	
τ=0.2		RMSE of τ	0.056	0.061	0.088	0.164	0.056	
β_1=2		RMSE of β_1	0.050	0.057	0.059	0.056	0.050	
β_2=1		RMSE of β_2	0.048	0.049	0.051	0.051	0.048	
		Loss	3.009	42.180	63.024	76.411	3.009	3.609
		MS accuracy	1.000	0.000	0.000	0.000		
		MA weights	0.924	0.043	0.022	0.011		
			W_1	W_2	W_3	W_4	MS	MA
α=−0.2	n=169	RMSE of α	0.035	0.054	0.072	0.153	0.037	
τ=−0.2		RMSE of τ	0.084	0.093	0.128	0.244	0.084	
β_1=2		RMSE of β_1	0.073	0.076	0.089	0.087	0.073	
β_2=1		RMSE of β_2	0.078	0.109	0.115	0.139	0.078	
		Loss	3.102	22.662	30.673	35.482	3.503	3.906
		MS accuracy	0.982	0.015	0.001	0.002		
		MA weights	0.882	0.061	0.023	0.034		
α=−0.2	n=400	RMSE of α	0.023	0.034	0.047	0.072	0.023	
τ=−0.2		RMSE of τ	0.056	0.062	0.089	0.191	0.056	
β_1=2		RMSE of β_1	0.049	0.054	0.054	0.052	0.049	
β_2=1		RMSE of β_2	0.056	0.056	0.058	0.060	0.056	
		Loss	3.273	45.621	64.126	71.918	3.273	3.956
		MS accuracy	1.000	0.000	0.000	0.000		
		MA weights	0.927	0.039	0.014	0.020		
			W_1	W_2	W_3	W_4	MS	MA
α=1.2	n=169	RMSE of α	0.028	0.858	5.851	5.232	0.057	
τ=1.2		RMSE of τ	0.082	0.715	5.779	5.262	0.084	
β_1=2		RMSE of β_1	0.056	0.965	1.069	0.920	0.090	
β_2=1		RMSE of β_2	0.054	0.610	0.597	0.508	0.070	
		Loss	6.928	1575.414	1802.360	1755.770	14.316	45.112
		MS accuracy	0.996	0.003	0.001	0.000		
		MA weights	0.915	0.043	0.022	0.020		
α=1.2	n=400	RMSE of α	0.020	0.449	3.939	4.359	0.020	
τ=1.2		RMSE of τ	0.053	0.379	4.041	4.406	0.053	
β_1=2		RMSE of β_1	0.041	0.844	0.763	0.617	0.041	
β_2=1		RMSE of β_2	0.032	0.440	0.410	0.368	0.032	
		Loss	8.719	3049.622	3430.742	3831.036	8.719	40.668
		MS accuracy	1.000	0.000	0.000	0.000		
		MA weights	0.940	0.027	0.019	0.014		
			W_1	W_2	W_3	W_4	MS	MA
α=−1.2	n=169	RMSE of α	0.030	0.664	6.880	11.470	0.614	
τ=−1.2		RMSE of τ	0.086	0.399	6.673	11.590	0.703	
β_1=2		RMSE of β_1	0.064	0.974	1.204	1.419	0.139	
β_2=1		RMSE of β_2	0.048	0.449	0.621	0.762	0.076	
		Loss	8.593	1627.103	1838.773	2096.612	20.260	57.258
		MS accuracy	0.995	0.001	0.002	0.002		
		MA weights	0.901	0.029	0.015	0.054		
α=−1.2	n=400	RMSE of α	0.020	0.423	8.093	11.373	0.020	
τ=−1.2		RMSE of τ	0.055	0.334	7.750	11.166	0.055	
β_1=2		RMSE of β_1	0.041	0.846	1.314	1.306	0.041	
β_2=1		RMSE of β_2	0.036	0.445	0.716	0.693	0.036	
		Loss	9.142	3197.465	4508.588	4673.683	9.142	54.185
		MS accuracy	1.000	0.000	0.000	0.000		
		MA weights	0.936	0.021	0.012	0.031		

Table 4: Simulation results when W_1 is the true matrix and error terms are heteroskedastic

			W_1	W_2	W_3	W_4	MS	MA
α=0.2	n=169	RMSE of α	0.035	0.051	0.069	0.093	0.036	
τ=0.2		RMSE of τ	0.086	0.096	0.138	0.207	0.086	
β_1=2		RMSE of β_1	0.099	0.102	0.102	0.107	0.099	
β_2=1		RMSE of β_2	0.085	0.088	0.093	0.097	0.085	
		Loss	4.322	22.655	30.453	35.390	4.714	5.015
		MS accuracy	0.980	0.020	0.000	0.000		
		MA weights	0.874	0.067	0.035	0.023		
α=0.2	n=400	RMSE of α	0.022	0.030	0.049	0.063	0.022	
τ=0.2		RMSE of τ	0.055	0.059	0.090	0.171	0.055	
β_1=2		RMSE of β_1	0.063	0.069	0.071	0.069	0.063	
β_2=1		RMSE of β_2	0.048	0.049	0.052	0.052	0.048	
		Loss	3.615	42.744	63.779	77.146	3.647	4.170
		MS accuracy	0.999	0.001	0.000	0.000		
		MA weights	0.924	0.041	0.026	0.009		
			W_1	W_2	W_3	W_4	MS	MA
α=−0.2	n=169	RMSE of α	0.035	0.054	0.076	0.154	0.037	
τ=−0.2		RMSE of τ	0.087	0.096	0.126	0.241	0.087	
β_1=2		RMSE of β_1	0.098	0.099	0.111	0.109	0.098	
β_2=1		RMSE of β_2	0.077	0.112	0.120	0.143	0.077	
		Loss	3.968	23.513	31.682	36.454	4.532	4.890
		MS accuracy	0.975	0.022	0.001	0.002		
		MA weights	0.870	0.070	0.024	0.036		
α=−0.2	n=400	RMSE of α	0.024	0.035	0.052	0.075	0.024	
τ=−0.2		RMSE of τ	0.057	0.063	0.090	0.194	0.057	
β_1=2		RMSE of β_1	0.068	0.071	0.071	0.070	0.068	
β_2=1		RMSE of β_2	0.056	0.057	0.060	0.064	0.056	
		Loss	4.438	46.739	65.444	73.118	4.476	5.100
		MS accuracy	0.999	0.001	0.000	0.000		
		MA weights	0.926	0.035	0.018	0.021		
			W_1	W_2	W_3	W_4	MS	MA
α=1.2	n=169	RMSE of α	0.034	0.902	6.636	6.885	0.052	
τ=1.2		RMSE of τ	0.086	0.940	6.451	6.973	0.099	
β_1=2		RMSE of β_1	0.077	0.906	1.153	0.901	0.108	
β_2=1		RMSE of β_2	0.054	0.338	0.597	0.527	0.062	
		Loss	12.911	1414.080	1877.172	1964.258	22.366	54.711
		MS accuracy	0.994	0.005	0.000	0.001		
		MA weights	0.907	0.036	0.024	0.034		
α=1.2	n=400	RMSE of α	0.020	0.409	3.200	7.221	0.022	
τ=1.2		RMSE of τ	0.055	0.317	3.330	7.273	0.063	
β_1=2		RMSE of β_1	0.049	0.853	0.728	0.869	0.049	
β_2=1		RMSE of β_2	0.031	0.402	0.382	0.467	0.035	
		Loss	10.173	3260.509	3574.774	4340.654	13.610	46.912
		MS accuracy	0.999	0.000	0.001	0.000		
		MA weights	0.943	0.024	0.015	0.018		
			W_1	W_2	W_3	W_4	MS	MA
α=−1.2	n=169	RMSE of α	0.032	0.694	8.035	10.733	0.086	
τ=−1.2		RMSE of τ	0.084	0.404	7.901	10.504	0.214	
β_1=2		RMSE of β_1	0.074	0.912	1.218	1.270	0.107	
β_2=1		RMSE of β_2	0.049	0.644	0.775	0.883	0.104	
		Loss	8.774	1410.008	1706.016	1854.547	21.983	57.736
		MS accuracy	0.993	0.002	0.000	0.005		
		MA weights	0.895	0.025	0.018	0.062		
α=−1.2	n=400	RMSE of α	0.021	0.425	8.314	11.379	0.021	
τ=−1.2		RMSE of τ	0.056	0.357	7.983	11.100	0.056	
β_1=2		RMSE of β_1	0.048	0.845	1.296	1.270	0.048	
β_2=1		RMSE of β_2	0.036	0.447	0.703	0.674	0.036	
		Loss	10.699	3192.960	4480.797	4622.047	10.699	60.891
		MS accuracy	1.000	0.000	0.000	0.000		
		MA weights	0.926	0.024	0.015	0.035		

Table 5: Simulation results when $W_2 + W_4$ is the true matrix and error terms are normal

			W_1	W_2	W_3	W_4	MS	MA
α=0.2	n=169	RMSE of α	0.087	0.079	0.151	0.239	0.130	
τ=0.2		RMSE of τ	0.173	0.089	0.138	0.359	0.148	
β_1=2		RMSE of β_1	0.126	0.077	0.078	0.090	0.077	
β_2=1		RMSE of β_2	0.094	0.081	0.084	0.084	0.081	
		Loss	25.059	6.213	15.484	16.662	8.935	5.733
		MS accuracy	0.014	0.736	0.118	0.132		
		MA weights	0.044	0.577	0.123	0.257		
α=0.2	n=400	RMSE of α	0.079	0.065	0.143	0.231	0.079	
τ=0.2		RMSE of τ	0.161	0.062	0.112	0.320	0.081	
β_1=2		RMSE of β_1	0.112	0.053	0.054	0.055	0.054	
β_2=1		RMSE of β_2	0.073	0.053	0.057	0.057	0.053	
		Loss	45.843	9.097	30.670	32.313	10.615	7.490
		MS accuracy	0.002	0.929	0.047	0.022		
		MA weights	0.025	0.659	0.101	0.215		
			W_1	W_2	W_3	W_4	MS	MA
α=−0.2	n=169	RMSE of α	0.072	0.068	0.139	0.203	0.149	
τ=−0.2		RMSE of τ	0.221	0.127	0.166	0.222	0.141	
β_1=2		RMSE of β_1	0.115	0.079	0.090	0.090	0.083	
β_2=1		RMSE of β_2	0.085	0.080	0.086	0.099	0.083	
		Loss	25.247	9.137	12.085	19.541	11.501	7.691
		MS accuracy	0.033	0.574	0.264	0.129		
		MA weights	0.065	0.539	0.183	0.213		
α=−0.2	n=400	RMSE of α	0.070	0.067	0.120	0.179	0.127	
τ=−0.2		RMSE of τ	0.217	0.096	0.133	0.142	0.103	
β_1=2		RMSE of β_1	0.118	0.051	0.051	0.058	0.052	
β_2=1		RMSE of β_2	0.060	0.051	0.053	0.074	0.056	
		Loss	65.561	18.298	30.144	34.524	22.219	11.016
		MS accuracy	0.001	0.701	0.130	0.168		
		MA weights	0.016	0.588	0.086	0.311		
			W_1	W_2	W_3	W_4	MS	MA
α=1.2	n=169	RMSE of α	0.878	0.419	1.203	1.460	0.513	
τ=1.2		RMSE of τ	0.470	0.819	0.887	0.562	0.812	
β_1=2		RMSE of β_1	1.356	0.199	0.901	0.539	0.347	
β_2=1		RMSE of β_2	0.644	0.171	0.675	0.384	0.240	
		Loss	2924.891	307.403	7413.951	1180.077	469.299	427.650
		MS accuracy	0.048	0.911	0.009	0.032		
		MA weights	0.096	0.732	0.019	0.152		
α=1.2	n=400	RMSE of α	0.938	0.186	1.198	1.682	0.206	
τ=1.2		RMSE of τ	0.578	0.501	0.655	0.566	0.502	
β_1=2		RMSE of β_1	1.091	0.191	0.859	0.589	0.197	
β_2=1		RMSE of β_2	0.610	0.075	0.756	0.319	0.080	
		Loss	7844.534	551.384	27743.911	3853.553	576.453	637.704
		MS accuracy	0.002	0.994	0.001	0.003		
		MA weights	0.030	0.828	0.015	0.128		
			W_1	W_2	W_3	W_4	MS	MA
α=−1.2	n=169	RMSE of α	0.898	0.445	0.219	1.111	0.989	
τ=−1.2		RMSE of τ	0.989	0.295	0.856	1.045	0.916	
β_1=2		RMSE of β_1	1.345	0.266	0.286	0.486	0.599	
β_2=1		RMSE of β_2	0.561	0.233	0.287	0.451	0.427	
		Loss	9263.240	2843.288	5380.361	5009.639	4957.023	4512.994
		MS accuracy	0.097	0.319	0.102	0.482		
		MA weights	0.118	0.344	0.082	0.456		
α=−1.2	n=400	RMSE of α	1.019	0.304	0.119	0.763	0.649	
τ=−1.2		RMSE of τ	1.248	0.396	0.728	0.711	0.611	
β_1=2		RMSE of β_1	1.349	0.120	0.141	0.508	0.467	
β_2=1		RMSE of β_2	0.658	0.066	0.090	0.368	0.311	
		Loss	14333.386	4178.053	5869.305	5241.881	4334.436	3802.430
		MS accuracy	0.021	0.423	0.120	0.436		
		MA weights	0.044	0.429	0.106	0.421		

7 An empirical illustration

Ertur and Koch (2007) explore the impact of spillover effects by incorporating technological interdependence into a growth model. They estimate a SAR model and find significant spatial externalities. In this empirical illustration, we discuss the choice of spatial weights matrix for the MESS version of their model.

Ertur and Koch (2007) use the following SAR specification:

$$\ln y = \lambda W \ln y + \beta_0 + \beta_1 \ln s + \beta_2 \ln(\mathrm{n} + 0.05 l_n) + \beta_3 W \ln s + \beta_4 W \ln(\mathrm{n} + 0.05 l_n) + \epsilon, \quad (7.1)$$

where y is the $n \times 1$ vector of the output per-worker, s is the $n \times 1$ vector of fraction of savings, n is the $n \times 1$ vector of exogenous growth rate of labor, l_n is the $n \times 1$ vector of ones, and ϵ is the $n \times 1$ vector of disturbance terms. The sum of the annual rate of depreciation of physical capital and the balanced growth rate of capital-output ratio is set to 0.05, which is a common assumption in the economic growth literature. Ertur and Koch (2007) consider two spatial weights matrices (i) $W_1^\kappa = (w_{1ij}^\kappa)$ and (ii) $W_2^\kappa = (w_{2ij}^\kappa)$, whose elements are specified as

$$w_{1ij}^\kappa = \begin{cases} 0 & \text{if } i = j, \\ d_{ij}^{-\kappa} & \text{if } i \neq j, \end{cases} \qquad w_{2ij}^\kappa = \begin{cases} 0 & \text{if } i = j, \\ e^{-\kappa\, d_{ij}} & \text{if } i \neq j, \end{cases} \quad (7.2)$$

where d_{ij} is the great-circle distance between country capitals, and κ is a constant that controls the rate of decline in the magnitude of the weights when d_{ij} increases. Both weights matrices are row normalized. Ertur and Koch (2007) chooses $\kappa = 2$ (i.e., $W_1^{\kappa=2}$ and $W_2^{\kappa=2}$), and show that there exists significant spillover effects in the output per-worker.

To implement our model selection and model averaging procedures, we generate six candidates for the spatial weights matrices, corresponding to $\kappa \in \{2, 1, 0.5\}$. We consider a cross-sectional data set of 91 countries, and estimate the following MESS version of (7.1).

$$e^{\alpha W} \ln y = \delta_0 + \delta_1 \ln s + \delta_2 \ln(\mathrm{n} + 0.05 l_n) + \delta_3 W \ln s + \delta_4 W \ln(\mathrm{n} + 0.05 l_n) + \nu, \quad (7.3)$$

where ν is the $n \times 1$ vector of disturbance terms. We compute our criterion function under each of the six candidate weights matrices. We also compute the MA weights for the first three, the last three, and all six spatial weights matrices as a group, respectively. We report the coefficient estimates of (7.3), and those reported in Ertur and Koch (2007) from the estimation of (7.1).

Table 6 reports the estimation results.[11] When we use $W_1^{\kappa=2}$, the estimate of λ is 0.74 under the SAR, and the estimate of α is -0.864 under the MESS.[12] When the spatial weights matrix is $W_2^{\kappa=2}$, these estimates are 0.658 and -0.744 in the SAR and MESS, respectively. When we use $W_1^{\kappa=2}$

[11] We do not report the standard errors (or the confidence intervals) in this table, because the large sample distribution of the MS and MA estimators are not standard (Leeb and Pötscher, 2005). Also, our aim here is on illustrating the proposed selection and averaging schemes, not on statistical inference.

[12] Note that, when the spatial weights matrix is row normalized, the spatial coefficients α in the MESS and λ in the SAR have the following relation: $\lambda = 1 - e^\alpha$ (Debarsy et al., 2015).

and $W_2^{\kappa=2}$, each specification reports similar estimates for the other parameters, which can be seen by comparing column (1) with (2), and column (3) with (6). In the case of $W_1^{\kappa=1}$, $W_1^{\kappa=0.5}$, $W_2^{\kappa=1}$ and $W_2^{\kappa=0.5}$, the estimates reported under MESS varies as we are using different spatial weights matrices. The results in the table indicate that the criterion function (denoted by "Criterion") associated with $W_2^{\kappa=1}$ in column (7) obtains the smallest value out of all 6 candidates. The last two rows in the table show our MA results. If we compute the MA weights for the six candidates together, $W_2^{\kappa=1}$ has the model weight of 1.000, which means that all weights are placed on $W_2^{\kappa=1}$.[13] This result is consistent with our MS result based on the value of the criterion function. If we compute the MA weights for W_1 group (columns (3) to (5)) and W_2 group (columns (6) to (8)), separately, we find that $W_1^{\kappa=1}$ receives the largest MA weight (0.657) in the first group, and $W_2^{\kappa=1}$ receives the largest MA weight (1.000) in the second group. Overall, our MS and MA results show that $W_2^{\kappa=1}$ can be the optimal spatial weights matrix out of the six candidates for this application (note that the optimality is defined here in the sense of Theorem 1).

The parameter estimates in Table 6 are not directly comparable because of the spatial lags of the endogenous and exogenous variables in the model (Arbia et al., 2020; LeSage and Pace, 2009). In the context of (7.3), the marginal effects of exogenous variables are $\partial \ln y / \partial \ln s = e^{-\alpha W}(\delta_1 I_n + \delta_3 W)$ and $\partial \ln y / \partial \ln(n + 0.05 l_n) = e^{-\alpha W}(\delta_2 I_n + \delta_4 W)$. To ease the interpretation of these marginal effects, LeSage and Pace (2009) suggest three impact measures: (i) the average direct effect, (ii) the average total effect and (iii) the average indirect effect. The average direct effect of $\ln s$ is defined by $\mathrm{tr}\left(e^{-\alpha W}(\delta_1 I_n + \delta_3 W)\right)/n$, the average total effect by $l_n'\left(e^{-\alpha W}(\delta_1 I_n + \delta_3 W)\right) l_n / n$, and the average indirect effect by $l_n'\left(e^{-\alpha W}(\delta_1 I_n + \delta_3 W)\right) l_n / n - \mathrm{tr}\left(e^{-\alpha W}(\delta_1 I_n + \delta_3 W)\right)/n$. These summary measures can be defined similarly for $\ln(n + 0.05 l_n)$. Table 7 reports the impact measure results for our empirical illustration. The results in this table indicate that the impact measure estimates vary across different spatial weight matrices. It is instructive to compare the impact measure estimates reported in the seventh column under the optimal $W_2^{\kappa=1}$ with those reported in the first two columns. Under our optimal $W_2^{\kappa=1}$, we obtain relatively smaller average direct effect of $\ln s$ than those reported in the first two columns. On the other hand, we obtain relatively larger average total and indirect effects of $\ln s$ than those reported in the first two columns. In the case of $\ln(n + 0.05 l_n)$, all impact measures under $W_2^{\kappa=1}$ are relatively smaller in magnitude than those reported in the first two columns.

[13] The MA estimator using a Mallows type criterion function can yield sparse weights vectors in finite samples for both underfitted and overfitted models. See Feng et al. (2020) for a detailed discussion on the sparsity of Mallows model averaging estimator.

Table 6: Estimation results under different spatial weights matrices

	SAR		MESS					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$W_1^{\kappa=2}$	$W_2^{\kappa=2}$	$W_1^{\kappa=2}$	$W_1^{\kappa=1}$	$W_1^{\kappa=0.5}$	$W_2^{\kappa=2}$	$W_2^{\kappa=1}$	$W_2^{\kappa=0.5}$
Constant	0.988	0.530	2.003	1.652	13.490	2.371	2.902	3.465
$\ln s$	0.825	0.792	0.960	0.960	0.867	0.904	0.764	0.776
$\ln(\mathrm{n}+0.05)$	−1.498	−1.451	−1.745	−1.641	−1.307	−1.648	−1.300	−1.299
$W \ln s$	−0.322	−0.372	−0.284	−0.032	2.681	−0.221	0.060	0.301
$W \ln(\mathrm{n}+0.05)$	0.571	0.137	0.511	0.526	3.373	0.364	0.291	0.658
$W \ln y$	0.740	0.658	−0.864	−1.240	−2.597	−0.744	−0.865	−1.146
Criterion			51.510	51.084	157.264	50.746	44.683	49.366
6 weights			0.000	0.000	0.000	0.000	1.000	0.000
3 weights			0.343	0.657	0.000	0.000	1.000	0.000

Table 7: Impact measures under different spatial weights matrices

	SAR		MESS					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	$W_1^{\kappa=2}$	$W_2^{\kappa=2}$	$W_1^{\kappa=2}$	$W_1^{\kappa=1}$	$W_1^{\kappa=0.5}$	$W_2^{\kappa=2}$	$W_2^{\kappa=1}$	$W_2^{\kappa=0.5}$
	Average direct effects							
$\ln s$	0.908	0.940	0.981	0.998	1.362	0.925	0.814	0.834
$\ln(\mathrm{n}+0.05)$	−1.652	−1.712	−1.786	−1.670	−0.971	−1.695	−1.331	−1.298
	Average total effects							
$\ln s$	1.935	1.935	1.602	3.207	47.625	1.438	1.959	3.389
$\ln(\mathrm{n}+0.05)$	−3.565	−3.565	−2.928	−3.853	27.744	−2.702	−2.396	−2.017
	Average indirect effects							
$\ln s$	1.027	0.995	0.621	2.210	46.264	0.513	1.145	2.555
$\ln(\mathrm{n}+0.05)$	−1.913	−1.853	−1.142	−2.183	28.715	−1.007	−1.066	−0.719

8 Conclusion

In this paper, we focused on a specification problem in terms of choosing spatial weights matrices for the MESS models. We proposed a model selection scheme using a Mallows type criterion function and showed that the resulting MS estimator is asymptotically optimal. We also showed that when the data generating process involves spatial effects, the MS estimator chooses the true weights matrix with probability approaching one in large samples. We further proposed a model averaging scheme that compromises across a set of candidate models. The resulting MA estimator is also asymptotically optimal and can provide assurance against selecting the incorrect model in finite samples. Our extensive Monte Carlo simulations showed that the proposed MS and MA estimators perform satisfactorily in finite samples. In an illustration from the empirical growth literature, we revisit a model of spillover effects and show how the proposed MS and MA methods can be helpful in sensitivity or robustness checks. These methods thus are useful for researchers who want to select an optimal spatial weights matrix out of a set of candidates and construct an optimal weights of these candidates in the MESS models. We only considered the first order and higher-order MESS models with cross-sectional data. In future studies, our approach can be extended to the panel data versions of the MESS with individual or interactive fixed effects. It will also be interesting to consider these procedures for the heterogeneous coefficients MESS models. In our analysis, we assumed that the set of spatial weights matrices contains only the exogenous candidates. It will be interesting to extend the MS and MA procedures to the MESS models with endogenous spatial weights matrices. We leave these extensions for future studies.

Appendix

A Some useful lemmas

In this section, we provide some lemmas for easy reference. The first lemma is the well-known Stein's lemma, and can be found in Stein (1981), the second lemma in, e.g., Gao et al. (2019) and Zhang (2021), and the third and fourth lemma in Debarsy et al. (2015). We use the first lemma to motivate the derivation of our criterion function. The second lemma is used in the proofs of Theorems 1 and 3. The third lemma shows that certain matrices are bounded in both row and column sum norms, and the fourth lemma provides the probability orders of certain terms.

Lemma A.1. *Let Z be a random variable such that $Z \sim N(0,1)$, and let $f: R \to R$ be an indefinite integral of the Lebesgue measurable function $f^{'}$. Suppose further that $\mathbb{E}\big|f^{'}(Z)\big| < \infty$. Then $\mathbb{E}\big(f^{'}(Z)\big) = \mathbb{E}\big(Zf(Z)\big)$.*

Lemma A.2. *Let $\tilde{s} = \text{argmin}_{s\in\{1,\ldots,S\}}\,(L_s + a_s + b)$, where a_s is a term related to s and b is a term unrelated to s. If $\sup_{s\in\{1,\ldots,S\}} \frac{|a_s|}{R_s} = o_p(1)$ and $\sup_{s\in\{1,\ldots,S\}} \left|\frac{L_s - R_s}{R_s}\right| = o_p(1)$, and there exists a positive constant c and a positive integer N such that when $n \geq N, \inf_{s\in S} R_s \geq c > 0$ almost surely, then $\frac{L_{\tilde{s}}}{\inf_{s\in\{1,\ldots,S\}} L_s} \to 1$ in probability.*

Lemma A.3. *Assume that Assumptions 1 and 2 hold. Then, $e^{\tau M}$ and $e^{\alpha W}$ are bounded in both row and column sum norms uniformly in τ and α, respectively.*

Lemma A.4. *Assume that Assumption 4 holds. Let A be an $n \times n$ matrix that is bounded in both row and column sum norms, and C be an $n \times p$ matrix whose elements are uniformly bounded. Then,*

1. *$\epsilon^{'}A\epsilon = O_p(n)$, $\mathbb{E}(\epsilon^{'}A\epsilon) = O(n)$, $n^{-1}\epsilon^{'}A\epsilon = n^{-1}\mathbb{E}(\epsilon^{'}A\epsilon) + o_p(1)$, and $n^{-1/2}C^{'}A\epsilon = O_p(1)$.*
2. *$\|e^{\tilde{c}A} - I_n\|_\infty = o_p(1)$ and $\|e^{\tilde{c}A} - I_n\|_1 = o_p(1)$, where $\tilde{c} = o_p(1)$.*

B Derivation of (3.6) and (3.7)

For a given (α_s, τ_s) value, the first order conditions of (2.3) with respect to β and σ^2 yield

$$\widehat{\beta}_s = \left(X' e^{\tau_s M_s'} e^{\tau_s M_s} X\right)^{-1} X^{'} e^{\tau_s M_s^{'}} e^{\tau_s M_s} e^{\alpha_s W_s} y, \tag{B.1}$$

$$\widehat{\sigma}_s^2 = \frac{1}{n}\left\|e^{\tau_s M_s}(e^{\alpha_s W_s} y - X\widehat{\beta}_s)\right\|^2. \tag{B.2}$$

Substituting (B.1) and (B.2) into (2.3), we obtain the following concentrated log-likelihood function under sth model:

$$\ell_s^c = -\ln\left\|G_s e^{\tau W_s} e^{\alpha_s W_s} y\right\|^2,$$

where we ignore the constant terms. Then, the conditions $\frac{\partial \ell_s^c(\widehat{\alpha}_s,\widehat{\tau}_s)}{\partial \widehat{\alpha}_s} = 0$ and $\frac{\partial \ell_s^c(\widehat{\alpha}_s,\widehat{\tau}_s)}{\partial \widehat{\tau}_s} = 0$ imply that

$$-\left\|\widehat{G}_s e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s} y\right\|^{-2} (y' e^{\widehat{\alpha}_s W_s'} e^{\widehat{\tau}_s M_s'} \widehat{G}_s e^{\widehat{\tau}_s M_s} W_s e^{\widehat{\alpha}_s W_s} y) = 0, \tag{B.3}$$

$$-\left\|\widehat{G}_s e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s} y\right\|^{-2} \left(y' e^{\widehat{\alpha}_s W_s'} e^{\widehat{\tau}_s M_s'} \left(2\widehat{G}_s M_s + \frac{\partial \widehat{G}_s}{\partial \widehat{\tau}_s}\right) e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s} y\right) = 0. \tag{B.4}$$

Taking the derivative of (B.3) with respect to y for a given $\widehat{\tau}_s$, and the derivative of (B.4) with respect to y for a given $\widehat{\alpha}_s$, we obtain respectively

$$-2a_1\left(A_1 + a_1\frac{\partial \widehat{\alpha}_s}{\partial y}\right)\left\|\widehat{G}_s e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s} y\right\|^{-2} + A_2 + A_3 + (a_2 + a_3)\frac{\partial \widehat{\alpha}_s}{\partial y} = 0, \tag{B.5}$$

$$-b_1\left(2B_1 + b_1\frac{\partial \widehat{\tau}_s}{\partial y}\right)\left\|\widehat{G}_s e^{\widehat{\tau}_s M_s} e^{\widehat{\alpha}_s W_s} y\right\|^{-2} + (2b_2 + b_3)\frac{\partial \widehat{\tau}_s}{\partial y} + 2B_2 = 0. \tag{B.6}$$

Then, (3.6) and (3.7) can be obtained by solving (B.5) and (B.6) for $\frac{\partial \widehat{\alpha}_s}{\partial y}$ and $\frac{\partial \widehat{\tau}_s}{\partial y}$.

C Expressions for MESS$(1,0)$ and MESS$(0,1)$

In this section, we consider two special cases of MESS$(1,1)$, namely, MESS$(1,0)$ and MESS$(0,1)$. Our results in Theorems 1–3 are also valid for these special cases. Here, our aim is to provide the required quantities for the model selection and averaging procedures. We start with the MESS$(1,0)$, which is given by

$$e^{\alpha W} y = X\beta + \epsilon, \tag{C.1}$$

with a reduced form of $y = e^{-\alpha W} X\beta + e^{-\alpha W}\epsilon$, implying that $\mu = \mathbb{E}(y) = e^{-\alpha W} X\beta$ and $\Omega = \sigma^2 e^{-\alpha W} e^{-\alpha W'}$. The quasi log-likelihood function for the sth model is then given by

$$\ell_s = -\frac{n}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}\left\|e^{\alpha W_s} y - X\beta\right\|^2. \tag{C.2}$$

For a given value $\widehat{\alpha}_s$, the estimator for μ_s is subsequently given by $\widehat{\mu}_s = e^{-\widehat{\alpha}_s W_s} X(X'X)^{-1}X' e^{\widehat{\alpha}_s W_s} y = \widetilde{P}_s y$, where $\widetilde{P}_s = e^{-\widehat{\alpha}_s W_s} X(X'X)^{-1}X' e^{\widehat{\alpha}_s W_s}$. Then, the feasible version of our selection criterion function takes the following form.

$$\widehat{C}_s = \left\|\widetilde{P}_s y - y\right\|^2 + 2\left(\operatorname{tr}\left(\widetilde{P}_s\widehat{\Omega}\right) + \frac{\partial \widehat{\alpha}_s}{\partial y'}\widehat{\Omega}\frac{\partial \widetilde{P}_s}{\partial \widehat{\alpha}_s} y\right), \tag{C.3}$$

where $\frac{\partial \widetilde{P}_s}{\partial \widehat{\alpha}_s} = -W_s\widetilde{P}_s + \widetilde{P}_s W_s$ and $\frac{\partial \widehat{\alpha}_s}{\partial y} = -\frac{2a_1 A_1 - \left\|\widehat{G}_s e^{\widehat{\alpha}_s W_s} y\right\|^2 (A_2+A_3)}{2a_1^2 - \left\|\widehat{G}_s e^{\widehat{\alpha}_s W_s} y\right\|^2 (a_2+a_3)}$, with

$$a_1 = y^{'} e^{\widehat{\alpha}_s W_s^{'}} \widehat{G}_s e^{\widehat{\alpha}_s W_s} W_s y, \quad a_2 = y^{'} W_s^{'} e^{\widehat{\alpha}_s W_s^{'}} \widehat{G}_s e^{\widehat{\alpha}_s W_s} W_s y, \quad a_3 = y^{'} e^{\widehat{\alpha}_s W_s^{'}} \widehat{G}_s e^{\widehat{\alpha}_s W_s} W_s^2 y,$$
$$A_1 = e^{\widehat{\alpha}_s W_s^{'}} \widehat{G}_s e^{\widehat{\alpha}_s W_s} y, \quad A_2 = e^{\widehat{\alpha}_s W_s^{'}} \widehat{G}_s e^{\widehat{\alpha}_s W_s} W_s y \quad \text{and} \quad A_3 = W_s^{'} e^{\widehat{\alpha}_s W_s^{'}} \widehat{G}_s e^{\widehat{\alpha}_s W_s} y,$$

and $G = I_n - P_X = I_n - X(X^{'}X)^{-1}X^{'}$. The model averaging estimator for μ is given by $\widehat{\mu}(w) = \widetilde{P}(w)y = \sum_{s=1}^{S} w_s \widetilde{P}_s y$ with a feasible weights choice criterion function stated as

$$\widehat{C}(w) = \left\|\widetilde{P}(w)y - y\right\|^2 + 2\left(\operatorname{tr}\left(\widetilde{P}(w)\widehat{\Omega}\right) + \sum_{s=1}^{S} w_s \frac{\partial \widehat{\alpha}_s}{\partial y^{'}} \widehat{\Omega} \frac{\partial \widetilde{P}_s}{\partial \widehat{\alpha}_s} y\right). \tag{C.4}$$

Next, we consider the MESS$(0,1)$, which is specified as

$$y = X\beta + u, \quad e^{\tau M} u = \epsilon, \tag{C.5}$$

with $\mu = \mathbb{E}(y) = X\beta$ and $\Omega = \sigma^2 e^{-\tau M} e^{-\tau M^{'}}$. The quasi log-likelihood for the sth model is given by

$$\ell_s = -\frac{n}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}\left\|e^{\tau M_s}(y - X\beta)\right\|^2. \tag{C.6}$$

The estimator for μ_s is given by $\widehat{\mu}_s = X(X^{'} e^{\widehat{\tau}_s M_s^{'}} e^{\widehat{\tau}_s M_s} X)^{-1} X^{'} e^{\widehat{\tau}_s M_s^{'}} e^{\widehat{\tau}_s M_s} y = \widetilde{P}_s y$, where $\widetilde{P}_s = X\left(X^{'} e^{\widehat{\tau}_s M_s^{'}} e^{\widehat{\tau}_s M_s} X\right)^{-1} X^{'} e^{\widehat{\tau}_s M_s^{'}} e^{\widehat{\tau}_s M_s}$. In this case, the feasible version of our selection criterion function is

$$C_s = \left\|\widetilde{P}_s y - y\right\|^2 + 2\left(\operatorname{tr}\left(\widetilde{P}_s \widehat{\Omega}\right) + \frac{\partial \widehat{\tau}_s}{\partial y^{'}} \widehat{\Omega} \frac{\partial \widetilde{P}_s}{\partial \widehat{\tau}_s} y\right), \tag{C.7}$$

where $\frac{\partial \widetilde{P}_s}{\partial \widehat{\tau}_s} = e^{-\widehat{\tau}_s M_s} \widehat{P}_s (M_s + M_s^{'})(I_n - \widehat{P}_s) e^{\widehat{\tau}_s M_s}$ and $\frac{\partial \widehat{\tau}_s}{\partial y} = -\frac{2b_1 B_1 - 2\left\|\widehat{G}_s e^{\widehat{\tau}_s M_s} y\right\|^2 B_2}{b_1^2 - \left\|\widehat{G}_s e^{\widehat{\tau}_s M_s} y\right\|^2 (2b_2 + b_3)}$, with

$$b_1 = y^{'} e^{\widehat{\tau}_s M_s^{'}} \left(2\widehat{G}_s M_s + \frac{\partial \widehat{G}_s}{\partial \widehat{\tau}_s}\right) e^{\widehat{\tau}_s M_s} y, \quad b_2 = y^{'} e^{\widehat{\tau}_s M_s^{'}} \left(2\widehat{G}_s M_s + \frac{\partial \widehat{G}_s}{\partial \widehat{\tau}_s}\right) M_s e^{\widehat{\tau}_s M_s} y,$$
$$b_3 = y^{'} e^{\widehat{\tau}_s M_s^{'}} \left(2\frac{\partial \widehat{G}_s}{\partial \widehat{\tau}_s} M_s + \frac{\partial^2 \widehat{G}_s}{\partial \widehat{\tau}_s^2}\right) e^{\widehat{\tau}_s M_s} y, \quad B_1 = e^{\widehat{\tau}_s M_s^{'}} \widehat{G}_s e^{\widehat{\tau}_s M_s} y \quad \text{and}$$
$$B_2 = e^{\widehat{\tau}_s M_s^{'}} \left(2\widehat{G}_s M_s + \frac{\partial \widehat{G}_s}{\partial \widehat{\tau}_s}\right) e^{\widehat{\tau}_s M_s} y.$$

The model averaging estimate for $\mu(w)$ is given by $\widehat{\mu}(w) = \widetilde{P}(w)y = \sum_{s=1}^{S} w_s \widetilde{P}_s y$ with the feasible weights choice criterion function,

$$\widehat{C}(w) = \left\| \widetilde{P}(w)y - y \right\|^2 + 2\left(\operatorname{tr}\left(\widetilde{P}(w)\widehat{\Omega} \right) + \sum_{s=1}^{S} w_s \frac{\partial \widehat{\tau}_s}{\partial y'} \widehat{\Omega} \frac{\partial \widetilde{P}_s}{\partial \widehat{\tau}_s} y \right). \tag{C.8}$$

D Expressions for the GMM approach

In this section, we derive the explicit expressions for $\frac{\partial \tilde{\alpha}_s}{\partial y'}$ and $\frac{\partial \tilde{\tau}_s}{\partial y'}$. According to Debarsy et al. (2015), the feasible best GMME is not available under an unknown form of heteroskedasticity. However, an optimal GMME is still available which uses the following vector of moment functions:

$$g(\gamma) = \frac{1}{n}\left(\epsilon'(\gamma) P_1^* \epsilon(\gamma), \epsilon'(\gamma) P_2^* \epsilon(\gamma), \epsilon'(\gamma) F^* \right)',$$

where $P_1^* = e^{\tau M} W e^{-\tau M} - \operatorname{Diag}(e^{\tau M} W e^{-\tau M}), P_2^* = M,\ F^* = (e^{\tau M} W X \beta, e^{\tau M} X)$, and $\operatorname{Diag}(A)$ denotes the $n \times n$ diagonal matrix consisting of the diagonal elements of A.

Let $\phi = (\alpha, \tau)'$, $P = e^{\tau M} X \left(X' e^{\tau M'} e^{\tau M} X \right)^{-1} X' e^{\tau M'}$ and $G = I_n - P$. Given ϕ, the estimate for β is $\widehat{\beta}(\phi) = \left(X' e^{\tau M'} e^{\tau M} X \right)^{-1} X' e^{\tau M'} e^{\tau M} e^{\alpha W} y$. The least square-type residual vector is then given by $\epsilon_x(\phi) = e^{\tau M}\left(e^{\alpha W} y - X\widehat{\beta}(\phi) \right) = G e^{\alpha W} e^{\tau M} y$. Then, the concentrated set of moment functions is given by

$$g(\phi) = \frac{1}{n}\left(\epsilon_x'(\phi) P_1^* \epsilon_x(\phi), \epsilon_x'(\phi) P_2^* \epsilon_x(\phi), \epsilon_x'(\phi) F^* \right)'.$$

The GMM objective function is given by $\Psi(\phi) = g'(\phi) V g(\phi)$, where V is an arbitrary weighting matrix. Note that an optimal GMME will be based on the weight matrix $\Omega_x = \operatorname{Var}(g(\phi))$. Consider the first-order conditions of the objective function $\frac{\partial \Psi(\phi)}{\partial \tilde{\alpha}_s} = 0$ and $\frac{\partial \Psi(\phi)}{\partial \tilde{\tau}_s} = 0$. The derivatives of these conditions with respect to y are

$$\frac{\partial \left(\partial \Psi(\phi) / \partial \tilde{\alpha}_s \right)}{\partial y'} + \frac{\partial \left(\partial \Psi(\phi) / \partial \tilde{\alpha}_s \right)}{\partial \tilde{\alpha}_s} \frac{\partial \tilde{\alpha}_s}{\partial y'} = 0, \tag{D.1}$$

$$\frac{\partial \left(\partial \Psi(\phi) / \partial \tilde{\tau}_s \right)}{\partial y'} + \frac{\partial \left(\partial \Psi(\phi) / \partial \tilde{\tau}_s \right)}{\partial \tilde{\tau}_s} \frac{\partial \tilde{\tau}_s}{\partial y'} = 0. \tag{D.2}$$

Then, using (D.1) and (D.2), we obtain

$$\frac{\partial \tilde{\alpha}_s}{\partial y'} = -\left(\frac{\partial \left(\partial \Psi(\phi) / \partial \tilde{\alpha}_s \right)}{\partial \tilde{\alpha}_s} \right)^{-1} \left(\frac{\partial \left(\partial \Psi(\phi) / \partial \tilde{\alpha}_s \right)}{\partial y'} \right), \tag{D.3}$$

$$\frac{\partial \tilde{\tau}_s}{\partial y'} = -\left(\frac{\partial \left(\partial \Psi(\phi) / \partial \tilde{\tau}_s \right)}{\partial \tilde{\tau}_s} \right)^{-1} \left(\frac{\partial \left(\partial \Psi(\phi) / \partial \tilde{\tau}_s \right)}{\partial y'} \right). \tag{D.4}$$

For the terms in (D.3) and (D.4), we can derive the following general expressions:

$$\frac{\partial \Psi(\phi)}{\partial \alpha} = g^{'}(\phi) V^{s} \frac{\partial g(\phi)}{\partial \alpha}, \tag{D.5}$$

$$\frac{\partial \left(\partial \Psi(\phi)/\partial \alpha\right)}{\partial \alpha} = \frac{\partial g^{'}(\phi)}{\partial \alpha} V^{s} \frac{\partial g(\phi)}{\partial \alpha} + g^{'}(\phi) V^{s} \frac{\partial g^{2}(\phi)}{\partial \alpha^{2}}, \tag{D.6}$$

$$\frac{\partial \left(\partial \Psi(\phi)/\partial \alpha\right)}{\partial y^{'}} = \frac{\partial g^{'}(\phi)}{\partial \alpha} V^{s} \frac{\partial g(\phi)}{\partial y^{'}} + g^{'}(\phi) V^{s} \frac{\partial g^{2}(\phi)}{\partial \alpha \partial y^{'}}, \tag{D.7}$$

$$\frac{\partial \Psi(\phi)}{\partial \tau} = g^{'}(\phi) V^{s} \frac{\partial g(\phi)}{\partial \tau}, \tag{D.8}$$

$$\frac{\partial \left(\partial \Psi(\phi)/\partial \tau\right)}{\partial \tau} = \frac{\partial g^{'}(\phi)}{\partial \tau} V^{s} \frac{\partial g(\phi)}{\partial \tau} + g^{'}(\phi) V^{s} \frac{\partial g^{2}(\phi)}{\partial \tau^{2}}, \tag{D.9}$$

$$\frac{\partial \left(\partial \Psi(\phi)/\partial \tau\right)}{\partial y^{'}} = \frac{\partial g^{'}(\phi)}{\partial \tau} V^{s} \frac{\partial g(\phi)}{\partial y^{'}} + g^{'}(\phi) V^{s} \frac{\partial g^{2}(\phi)}{\partial \tau \partial y^{'}}. \tag{D.10}$$

Using the vector of moment functions $g(\phi)$, the terms in (D.5) through (D.10) can be derived as

$$\frac{\partial g(\phi)}{\partial \alpha} = \begin{pmatrix} \epsilon_{x}^{'}(\phi) P_{1}^{*s} G e^{\tau M} e^{\alpha W} W y \\ \epsilon_{x}^{'}(\phi) P_{2}^{*s} G e^{\tau M} e^{\alpha W} W y \\ F^{*s^{'}} G e^{\tau M} e^{\alpha W} W y \end{pmatrix}, \tag{D.11}$$

$$\frac{\partial g(\phi)}{\partial \tau} = \begin{pmatrix} \epsilon_{x}^{'}(\phi) P_{1}^{*s} \frac{\partial \epsilon_{x}(\phi)}{\partial \tau} \\ \epsilon_{x}^{'}(\phi) P_{2}^{*s} \frac{\partial \epsilon_{x}(\phi)}{\partial \tau} \\ F^{*^{'}} G M e^{\tau M} e^{\alpha W} y \end{pmatrix}, \tag{D.12}$$

$$\frac{\partial g(\phi)}{\partial y^{'}} = \begin{pmatrix} \epsilon_{x}^{'}(\phi) P_{1}^{*s} G e^{\tau M} e^{\alpha W} \\ \epsilon_{x}^{'}(\phi) P_{2}^{*s} G e^{\tau M} e^{\alpha W} \\ F^{*s^{'}} G M e^{\tau M} e^{\alpha W} \end{pmatrix}, \tag{D.13}$$

$$\frac{\partial^{2} g(\phi)}{\partial \alpha \partial y^{'}} = \begin{pmatrix} y^{'} W^{'} e^{\alpha W^{'}} e^{\tau M^{'}} G P_{1}^{*s} G e^{\tau M} e^{\alpha W} + \epsilon_{x}^{'}(\phi) P_{1}^{*s} G e^{\tau M} e^{\alpha W} W \\ y^{'} W^{'} e^{\alpha W^{'}} e^{\tau M^{'}} G P_{2}^{*s} G e^{\tau M} e^{\alpha W} + \epsilon_{x}^{'}(\phi) P_{2}^{*s} G e^{\tau M} e^{\alpha W} W \\ F^{*s^{'}} G e^{\tau M} e^{\alpha W} W \end{pmatrix}, \tag{D.14}$$

$$\frac{\partial^{2} g(\phi)}{\partial \tau \partial y^{'}} = \begin{pmatrix} \frac{\partial \epsilon_{x}^{'}(\phi)}{\partial \tau} P_{1}^{*s} e^{\alpha W} e^{\tau M} G + \epsilon_{x}^{'}(\phi) P_{1}^{*s} \frac{\partial^{2} \epsilon_{x}(\phi)}{\partial \tau \partial y^{'}} \\ \frac{\partial \epsilon_{x}^{'}(\phi)}{\partial \tau} P_{2}^{*s} e^{\alpha W} e^{\tau M} G + \epsilon_{x}^{'}(\phi) P_{2}^{*s} \frac{\partial^{2} \epsilon_{x}(\phi)}{\partial \tau \partial y^{'}} \\ F^{*s^{'}} G M e^{\tau M} e^{\alpha W} \end{pmatrix}, \tag{D.15}$$

$$\frac{\partial^2 g(\phi)}{\partial \alpha^2} = \begin{pmatrix} y'W'e^{\alpha W'}e^{\tau M'}GP_1^{*s}Ge^{\tau M}e^{\alpha W}Wy + \epsilon_x'(\phi)P_1^{*s}Ge^{\tau M}e^{\alpha W}W^2y \\ y'W'e^{\alpha W'}e^{\tau M'}GP_2^{*s}Ge^{\tau M}e^{\alpha W}Wy + \epsilon_x'(\phi)P_2^{*s}Ge^{\tau M}e^{\alpha W}W^2y \\ F^{*s'}Ge^{\tau M}e^{\alpha W}W^2y \end{pmatrix}, \tag{D.16}$$

and

$$\frac{\partial^2 g(\phi)}{\partial \tau^2} = \begin{pmatrix} y'e^{\alpha W'}e^{\tau M'}M'GP_1^{*s}\frac{\partial \epsilon_x(\phi)}{\partial \tau} + \epsilon_x'(\phi)P_1^{*s}\left(\left(\frac{\partial G}{\partial \tau}M - \frac{\partial^2 P}{\partial \tau^2}\right)e^{\tau M}e^{\alpha W}y + \frac{\partial \epsilon_x^2(\phi)}{\partial \tau^2}\right) \\ y'e^{\alpha W'}e^{\tau M'}M'GP_2^{*s}\frac{\partial \epsilon_x(\phi)}{\partial \tau} + \epsilon_x'(\phi)P_2^{*s}\left(\left(\frac{\partial G}{\partial \tau}M - \frac{\partial^2 P}{\partial \tau^2}\right)e^{\tau M}e^{\alpha W}y + \frac{\partial \epsilon_x^2(\phi)}{\partial \tau^2}\right) \\ F^{*'}GM^2e^{\tau M}e^{\alpha W}y \end{pmatrix}, \tag{D.17}$$

where $\frac{\partial \epsilon_x(\phi)}{\partial \tau} = (GM - MP - PM^sP - PM')e^{\tau M}e^{\alpha W}y$, $\frac{\partial^2 \epsilon_x(\phi)}{\partial \tau \partial y'} = (GM - MP - PM^sP - PM')e^{\tau M}e^{\alpha W}$, $\frac{\partial \epsilon_x^2(\phi)}{\partial \tau^2} = \left(GM - MP - PM^sP - PM'\right)Me^{\tau M}e^{\alpha W}y$, $\frac{\partial G}{\partial \tau} = -MP - PM^sP - PM'$ and $\frac{\partial^2 P}{\partial \tau^2} = -\big(M(GM - MP - PM^sP - PM') + 2(GM - MP - PM^sP - PM')M^sP + (GM - MP - PM^sP - PM')M'\big)$. These expressions can be substituted back into (D.5) through (D.10), which in turn can be substituted into (D.3) and (D.4) to get the explicit expressions of the two derivatives.

E Proofs of Theorems

E.1 Proof of Theorem 1

First observe that the first term in the definition of $\widehat{C}_s$ in (3.8) satisfies the following equation:

$$\begin{aligned} \left\|\widetilde{P}_s y - y\right\|^2 &= \left\|\widetilde{P}_s y - \mu + \mu - y\right\|^2 \\ &= L_s + \|\tilde{\epsilon}\|^2 - 2\tilde{\epsilon}'(\widetilde{P}_s y - \mu) = L_s + \|\tilde{\epsilon}\|^2 - 2\tilde{\epsilon}'\left(\widetilde{P}_s(\mu + \tilde{\epsilon}) - \mu\right) \\ &= L_s + \|\tilde{\epsilon}\|^2 + 2\tilde{\epsilon}'\widetilde{H}_s\mu - 2\tilde{\epsilon}'\widetilde{P}_s\tilde{\epsilon}, \end{aligned} \tag{E.1}$$

where $\widetilde{H}_s = I_n - \widetilde{P}_s$.

Note here $\|\tilde{\epsilon}\|^2$ is irrelevant for s. Thus, the selected model $\widehat{s}$ also minimizes the following term:

$$L_s + 2\tilde{\epsilon}\widetilde{H}_s\mu + 2\left(\mathrm{tr}\left(\widetilde{P}_s\widehat{\Omega}\right) - \tilde{\epsilon}'\widetilde{P}_s\tilde{\epsilon}\right) + 2\left(\frac{\partial \widehat{\alpha}_s}{\partial y}\widehat{\Omega}\frac{\partial \widetilde{P}_s}{\partial \widehat{\alpha}_s}y + \frac{\partial \widehat{\tau}_s}{\partial y}\widehat{\Omega}\frac{\partial \widetilde{P}_s}{\partial \widehat{\tau}_s}y\right). \tag{E.2}$$

By Lemma A.2, we need to show the following results:

$$\sup_s R_s^{*-1}\left|\tilde{\epsilon}'\widetilde{H}_s\mu\right| = o_p(1), \tag{E.3}$$

$$\sup_s R_s^{*-1}\left|\mathrm{tr}(\widetilde{P}_s\widehat{\Omega}) - \tilde{\epsilon}'\widetilde{P}_s\tilde{\epsilon}\right| = o_p(1), \tag{E.4}$$

$$\sup_s R_s^{*-1}\left|\frac{\partial\widehat{\alpha}_s}{\partial y'}\widehat{\Omega}\frac{\partial\widetilde{P}_s}{\partial\widehat{\alpha}_s}y\right| = o_p(1), \tag{E.5}$$

$$\sup_s R_s^{*-1}\left|\frac{\partial\widehat{\tau}_s}{\partial y'}\widehat{\Omega}\frac{\partial\widetilde{P}_s}{\partial\widehat{\tau}_s}y\right| = o_p(1), \tag{E.6}$$

$$\sup_s R_s^{*-1}\left|L_s - R_s^*\right| = o_p(1). \tag{E.7}$$

For (E.3), first observe $R_s^{*-1}\left|\tilde{\epsilon}'\widetilde{H}_s\mu\right| \leq R_s^{*-1}\left|\tilde{\epsilon}'H_s^*\mu\right| + R_s^{*-1}\left|\tilde{\epsilon}'(P_s^* - \widetilde{P}_s)\mu\right|$, where $H_s^* = I_n - P_s^*$. For the first term, we have

$$\begin{aligned}
P\left(\sup_s R_s^{*-1}\left|\tilde{\epsilon}'H_s^*\mu\right| > \eta\right) &\leq \sum_{s=1}^{S}\frac{\mathbb{E}\left(\left(\tilde{\epsilon}'H_s^*\mu\right)^{2G}\right)}{R_s^{*2G}\eta^{2G}} \leq c_1\eta^{-2G}\sum_{s=1}^{S}\|H_s^*\mu\|^{2G}R_s^{*-2G} \\
&\leq c_1\eta^{-2G}\sum_{s=1}^{S}R_s^{*-G} = o(1),
\end{aligned}$$

for some constant terms c_1 and G, where the first inequality follows from the generalized Chebyshev's inequality, the second follows from Assumption 4 and Theorem 2 in Whittle (1960), the third follows from the fact that $R_s^* = \mathbb{E}\,\|\mu_s^* - \mu\|^2 = \mathrm{tr}(P_s^*\Omega P_s^{*'}) + \|H_s^*\mu\|^2 \geq \|H_s^*\mu\|^2$ and the last equality follows from Assumption 5. For the second term, using the Cauchy-Schwarz inequality and the bounds of Rayleigh quotient, we obtain

$$\begin{aligned}
\sup_s R_s^{*-1}\left|\tilde{\epsilon}'(P_s^* - \widetilde{P}_s)\mu\right| &\leq \sup_s R_s^{*-1}\|\tilde{\epsilon}\|\left\|(P_s^* - \widetilde{P}_s)\mu\right\| \leq \zeta_n^{-1}\|\tilde{\epsilon}\|\sup_s \gamma_{\max}(P_s^* - \widetilde{P}_s)\|\mu\| \\
&= o_p(1),
\end{aligned}$$

by Assumptions 6 and 8, and the fact that $\|\tilde{\epsilon}\| = O_p(n^{1/2})$, which is ensured by Lemma A.4.

For (E.4), first note that $\left|\mathrm{tr}(\widetilde{P}_s\widehat{\Omega}) - \tilde{\epsilon}'\widetilde{P}_s\tilde{\epsilon}\right| \leq \left|\mathrm{tr}(P_s^*\Omega) - \tilde{\epsilon}'P_s^*\tilde{\epsilon}\right| + \left|\mathrm{tr}\left(P_s^*(\widehat{\Omega} - \Omega)\right)\right| + \left|\mathrm{tr}\left((\widetilde{P}_s - P_s^*)\widehat{\Omega}\right)\right| + \left|\tilde{\epsilon}'(\widetilde{P}_s - P_s^*)\tilde{\epsilon}\right|$. Define $D_s = P_s^* e^{-\alpha W}e^{-\tau M}$. Then, for the first term, we have

$$\begin{aligned}
P\left(\sup_s R_s^{*-1}\left(\mathrm{tr}(P_s^*\Omega) - \tilde{\epsilon}'P_s^*\tilde{\epsilon}\right) > \eta\right) &\leq \sum_{s=1}^{S}\frac{\mathbb{E}\left(\left(\mathrm{tr}(P_s^*\Omega) - \tilde{\epsilon}'P_s^*\tilde{\epsilon}\right)^{2G}\right)}{R_s^{*2G}\eta^{2G}} \leq c_2\sum_{s=1}^{S}\frac{\left(\mathrm{tr}(D_sD_s')\right)^{G}}{R_s^{*2G}\eta^{2G}} \\
&\leq c_2'\sum_{s=1}^{S}R_s^{*-G} = o(1),
\end{aligned}$$

for some constants c_2 and $c_2^{'}$, where the first inequality follows from generalized Chebyshev's inequality, the second from Assumption 4 and Theorem 2 in Whittle (1960), the third from $\sigma^2 \text{tr}(D_s D_s^{'}) = \text{tr}(P_s^* \Omega P_s^{*'}) \leq R_s^*$ and the fourth equality from Assumption 5.

To find the order of the second term, $\sup_s R_s^{*-1} \left| \text{tr}\left(P_s^* (\widehat{\Omega} - \Omega) \right) \right|$, note that for two $n \times n$ matrices Δ_1 and Δ_2, we have $\gamma_{\max}(\Delta_1 \Delta_2) \leq \gamma_{\max}(\Delta_1)\gamma_{\max}(\Delta_2)$ and $\gamma_{\max}(\Delta_1 + \Delta_2) \leq \gamma_{\max}(\Delta_1) + \gamma_{\max}(\Delta_2)$ (Li, 1987). Since $\gamma_{\max}(\Delta_1) \leq \|\Delta_1\|$ for any matrix norm $\|\cdot\|$, and $\gamma_{\max}(e^{\tau_s^* M_s} X (X^{'} e^{\tau_s^* M_s^{'}} e^{\tau_s^* M_s} X)^{-1} X^{'} e^{\tau_s^* M_s^{'}}) = 1$, it follows from Lemma A.3 that

$$\sup_s \gamma_{\max}(P_s^*) \leq \sup_s \gamma_{\max}(e^{-\alpha_s^* W_s}) \sup_s \gamma_{\max}(e^{-\tau_s^* M_s}) \sup_s \gamma_{\max}(e^{\alpha_s^* W_s}) \sup_s \gamma_{\max}(e^{\tau_s^* M_s}) = O(1), \tag{E.8}$$

$$\sup_s \gamma_{\max}(H_s^*) \leq 1 + \sup_s \gamma_{\max}(P_s^*) = O(1). \tag{E.9}$$

Thus,

$$\begin{aligned}
\sup_s R_s^{*-1} \left| \text{tr}\left(P_s^* (\widehat{\Omega} - \Omega) \right) \right| &\leq \zeta_n^{-1} \sup_s \left(\gamma_{\max}(\widehat{\Omega} - \Omega) \gamma_{\max}(P_s^*) \text{rank}(P_s^*) \right) \\
&\leq k \zeta_n^{-1} \left(\gamma_{\max}(\widehat{\Omega}) + \gamma_{\max}(\Omega) \right) \sup_s \gamma_{\max}(P_s^*) = o_p(1),
\end{aligned}$$

by (E.8), Assumption 8 and Lemma A.3. For the third term, $\sup_s R_s^{*-1} \left| \text{tr}(\widetilde{P}_s - P_s^*)\widehat{\Omega} \right|$, we have

$$\begin{aligned}
\sup_s R_s^{*-1} \left| \text{tr}\left((\widetilde{P}_s - P_s^*)\widehat{\Omega} \right) \right| &\leq \zeta_n^{-1} \sup_s \left(\gamma_{\max}(\widetilde{P}_s - P_s^*) \gamma_{\max}(\widehat{\Omega}) \text{rank}(\widetilde{P}_s - P_s^*) \right) \\
&\leq 2k \zeta_n^{-1} \sup_s \gamma_{\max}(\widetilde{P}_s - P_s^*) \gamma_{\max}(\widehat{\Omega}) = o_p(1),
\end{aligned}$$

by Assumption 8 and Lemma A.3. Finally, for the fourth term, using the bounds of Rayleigh quotient, we obtain $\sup_s R_s^{-1*} \left| \tilde{\epsilon}^{'} (\widetilde{P}_s - P_s^*) \tilde{\epsilon} \right| \leq \zeta_n^{-1} \|\tilde{\epsilon}\|^2 \sup_s \gamma_{\max}(\widetilde{P}_s - P_s^*) = o_p(1)$ by Assumption 8 and Lemma A.4.

Next, we consider the results in (E.5) and (E.6). It is easy to see that $\sup_s R_s^{*-1} \left| \frac{\partial \widehat{\alpha}_s}{\partial y^{'}} \widehat{\Omega} \frac{\partial \widetilde{P}_s}{\partial \widehat{\alpha}_s} y \right| \leq \zeta_n^{-1} \sup_s \left| \frac{\partial \widehat{\alpha}_s}{\partial y^{'}} \widehat{\Omega} \frac{\partial \widetilde{P}_s}{\partial \widehat{\alpha}_s} y \right| = o_p(1)$ by Assumption 7. Similarly, we can obtain the result in (E.6).

Now consider the result in (E.7). First note that we can express $|L_s - R_s^*|$ as

$$
\begin{aligned}
|L_s - R_s^*| &= \left| \left\| \widetilde{P}_s y - \mu \right\|^2 - \operatorname{tr}(P_s^* \Omega P_s^{*'}) - \|H_s^* \mu\|^2 \right| \\
&= \left| \left\| (\widetilde{P}_s - P_s^*)\mu \right\|^2 + \left\| (\widetilde{P}_s - P_s^*)\tilde{\epsilon} \right\|^2 + \|P_s^* \tilde{\epsilon}\|^2 - \operatorname{tr}(P_s^* \Omega P_s^{*'}) + 2\mu^{'} (\widetilde{P}_s - P_s^*)^{'} (\widetilde{P}_s - P_s^*)\tilde{\epsilon} \right. \\
&\quad + 2\mu^{'} (\widetilde{P}_s - P_s^*)^{'} P_s^* \tilde{\epsilon} - 2\mu^{'} (\widetilde{P}_s - P_s^*)^{'} H_s^* \mu + 2\tilde{\epsilon}^{'} (\widetilde{P}_s - P_s^*)^{'} P_s^* \tilde{\epsilon} - 2\tilde{\epsilon}^{'} (\widetilde{P}_s - P_s^*)^{'} H_s^* \mu \\
&\quad \left. - 2\tilde{\epsilon}^{'} P_s^{*'} H_s^* \mu \right| \\
&\leq \left\| (\widetilde{P}_s - P_s^*)\mu \right\|^2 + \left\| (\widetilde{P}_s - P_s^*)\tilde{\epsilon} \right\|^2 + \left| \|P_s^* \tilde{\epsilon}\|^2 - \operatorname{tr}(P_s^* \Omega P_s^{*'}) \right| + 2 \left| \mu^{'} (\widetilde{P}_s - P_s^*)^{'} (\widetilde{P}_s - P_s^*)\tilde{\epsilon} \right| \\
&\quad + 2 \left| \mu^{'} (\widetilde{P}_s - P_s^*)^{'} P_s^* \tilde{\epsilon} \right| + 2 \left| \mu^{'} (\widetilde{P}_s - P_s^*)^{'} H_s^* \mu \right| + 2 \left| \tilde{\epsilon}^{'} (\widetilde{P}_s - P_s^*)^{'} P_s^* \tilde{\epsilon} \right| + 2 \left| \tilde{\epsilon}^{'} (\widetilde{P}_s - P_s^*)^{'} H_s^* \mu \right| \\
&\quad + 2 \left| \tilde{\epsilon}^{'} P_s^{*'} H_s^* \mu \right| .
\end{aligned}
$$

Thus, we need to prove the following results for showing that $\sup_s R_s^{*-1} |L_s - R_s^*| = o_p(1)$:

$$\sup_s R_s^{*-1} \left\| (\widetilde{P}_s - P_s^*)\mu \right\|^2 = o_p(1), \tag{E.10}$$

$$\sup_s R_s^{*-1} \left\| (\widetilde{P}_s - P_s^*)\tilde{\epsilon} \right\|^2 = o_p(1), \tag{E.11}$$

$$\sup_s R_s^{*-1} \left| \|P_s^* \tilde{\epsilon}\|^2 - \operatorname{tr}(P_s^* \Omega P_s^{*'}) \right| = o_p(1), \tag{E.12}$$

$$\sup_s R_s^{*-1} \left| \mu^{'} (\widetilde{P}_s - P_s^*)^{'} (\widetilde{P}_s - P_s^*)\tilde{\epsilon} \right| = o_p(1), \tag{E.13}$$

$$\sup_s R_s^{*-1} \left| \mu^{'} (\widetilde{P}_s - P_s^*)^{'} P_s^* \tilde{\epsilon} \right| = o_p(1), \tag{E.14}$$

$$\sup_s R_s^{*-1} \left| \mu^{'} (\widetilde{P}_s - P_s^*)^{'} H_s^* \mu \right| = o_p(1), \tag{E.15}$$

$$\sup_s R_s^{*-1} \left| \tilde{\epsilon}^{'} (\widetilde{P}_s - P_s^*)^{'} P_s^* \tilde{\epsilon} \right| = o_p(1), \tag{E.16}$$

$$\sup_s R_s^{*-1} \left| \tilde{\epsilon}' (\widetilde{P}_s - P_s^*)^{'} H_s^* \mu \right| = o_p(1), \tag{E.17}$$

$$\sup_s R_s^{*-1} \left| \tilde{\epsilon}^{'} P_s^{*'} H_s^* \mu \right| = o_p(1). \tag{E.18}$$

For the result in (E.10), it follows from the bounds of Rayleigh quotient, Assumptions 6 and 8 that

$$\sup_s R_s^{*-1} \left\| (\widetilde{P}_s - P_s^*)\mu \right\|^2 \leq \zeta_n^{-1} \|\mu\|^2 \sup \gamma_{\max}^2 (\widetilde{P}_s - P_s^*) = o_p(1).$$

The result in (E.11) can be obtained similarly. The proof for the result in (E.12) is similar to the proof of the result in (E.4). For the result in (E.13), we have $\sup_s R_s^{*-1} \left| \mu^{'} (\widetilde{P}_s - P_s^*)^{'} (\widetilde{P}_s - P_s^*)\tilde{\epsilon} \right| \leq \zeta_n^{-1} \|\mu\| \sup_s \gamma_{\max}^2 (\widetilde{P}_s - P_s^*) \|\tilde{\epsilon}\| = o_p(1)$ by Assumption 6 and 8 and Lemma A.4. The remaining results in (E.14)-(E.18) can be obtained similarly by using (E.8), (E.9), Assumptions 6 and 8. This completes the proof of Theorem 1.

E.2 Proof of Theorem 2

Since $\left\|\widetilde{P}_s y - y\right\|^2 = L_s + \|\tilde{\epsilon}\|^2 + 2\tilde{\epsilon}^{'}\widetilde{H}_s\mu - 2\tilde{\epsilon}^{'}\widetilde{P}_s\tilde{\epsilon}$, our criterion function can be expressed as

$$\begin{aligned}\widehat{C}_s &= L_s + \|\tilde{\epsilon}\|^2 + 2\tilde{\epsilon}\widetilde{H}_s\mu + 2\left(\operatorname{tr}\left(\widetilde{P}_s\widehat{\Omega}\right) - \tilde{\epsilon}^{'}\widetilde{P}_s\tilde{\epsilon}\right) + 2\left(\frac{\partial\widehat{\alpha}_s}{\partial y}\widehat{\Omega}\frac{\partial\widetilde{P}_s}{\partial\widehat{\alpha}_s}y + \frac{\partial\widehat{\tau}_s}{\partial y}\widehat{\Omega}\frac{\partial\widetilde{P}_s}{\partial\widehat{\tau}_s}y\right)\\ &= R_s^* + q_n,\end{aligned}$$

where $q_n = L_s - R_s^* + \|\tilde{\epsilon}\|^2 + 2\tilde{\epsilon}\widetilde{H}_s\mu + 2\left(\operatorname{tr}\left(\widetilde{P}_s\widehat{\Omega}\right) - \tilde{\epsilon}^{'}\widetilde{P}_s\tilde{\epsilon}\right) + 2\left(\frac{\partial\widehat{\alpha}_s}{\partial y}\widehat{\Omega}\frac{\partial\widetilde{P}_s}{\partial\widehat{\alpha}_s}y + \frac{\partial\widehat{\tau}_s}{\partial y}\widehat{\Omega}\frac{\partial\widetilde{P}_s}{\partial\widehat{\tau}_s}y\right)$. We will first show that $q_n/\zeta_n^* = o_p(1)$. It follows from Assumption 9 and Lemma A.4 that $\zeta_n^{*-1}\|\tilde{\epsilon}\|^2 = o(1/n)O_p(n) = o_p(1)$. It directly follows from Assumption 9 that $\sup_s\left|\frac{\partial\widehat{\alpha}_s}{\partial y^{'}}\widehat{\Omega}\frac{\partial\widetilde{P}_s}{\partial\widehat{\alpha}_s}y\right|/\zeta_n^* = o_p(1)$ and $\sup_s\left|\frac{\partial\widehat{\tau}_s}{\partial y^{'}}\widehat{\Omega}\frac{\partial\widetilde{P}_s}{\partial\widehat{\tau}_s}y\right|/\zeta_n^* = o_p(1)$. Next, we will show that $\left|\tilde{\epsilon}\widetilde{H}_s\mu\right|/\zeta_n^* = o_p(1)$. Note that $\zeta_n^{*-1}\left|\tilde{\epsilon}^{'}\widetilde{H}_s\mu\right| \leq \zeta_n^{*-1}\left|\tilde{\epsilon}^{'}H_s^*\mu\right| + \zeta_n^{*-1}\left|\tilde{\epsilon}^{'}(P_s^* - \widetilde{P}_s)\mu\right|$, where $H_s^* = I_n - P_s^*$. It follows from Chebyshev's inequality and the bounds of Rayleigh quotient that

$$\begin{aligned}P\left(\zeta_n^{*-1}\left|\tilde{\epsilon}^{'}H_s^*\mu\right| > \eta\right) &= P\left(\left|\tilde{\epsilon}^{'}H_s^*\mu\right| > \zeta_n^*\eta\right) \leq \eta^{-2}\zeta_n^{*-2}\mathbb{E}\left(\tilde{\epsilon}^{'}H_s^*\mu\right)^2 \\ &= \eta^{-2}\zeta_n^{*-2}\operatorname{tr}\left(\mu^{'}H_s^{*'}\Omega H_s^*\mu\right) \leq \eta^{-2}\zeta_n^{*-2}\|\mu\|^2\gamma_{\max}^2(H_s^*)\gamma_{\max}(\Omega) = o(1),\end{aligned} \tag{E.19}$$

by Assumption 9 and Lemma A.3. Similarly, we have

$$\begin{aligned}P\left(\zeta_n^{*-1}\left|\tilde{\epsilon}^{'}(P_s^* - \widetilde{P}_s)\mu\right| > \eta\right) &\leq \eta^{-2}\zeta_n^{*-2}\mathbb{E}\left(\tilde{\epsilon}^{'}(P_s^* - \widetilde{P}_s)\mu\right)^2 \\ &\leq \eta^{-2}\zeta_n^{*-2}\gamma_{\max}^2\left(P_s^* - \widetilde{P}_s\right)\gamma_{\max}(\Omega)\|\mu\|^2 = o_p(1),\end{aligned} \tag{E.20}$$

by Assumption 9 and Lemma A.3. Next, we consider $\zeta_n^{*-1}\left|\operatorname{tr}(\widetilde{P}_s\widehat{\Omega}) - \tilde{\epsilon}^{'}\widetilde{P}_s\tilde{\epsilon}\right|$. Note that $\left|\operatorname{tr}(\widetilde{P}_s\widehat{\Omega}) - \tilde{\epsilon}^{'}\widetilde{P}_s\tilde{\epsilon}\right| \leq \left|\operatorname{tr}(P_s^*\Omega) - \tilde{\epsilon}^{'}P_s^*\tilde{\epsilon}\right| + \left|\operatorname{tr}\left(P_s^*(\widehat{\Omega} - \Omega)\right)\right| + \left|\operatorname{tr}\left((\widetilde{P}_s - P_s^*)\widehat{\Omega}\right)\right| + \left|\tilde{\epsilon}^{'}(\widetilde{P}_s - P_s^*)\tilde{\epsilon}\right|$. Then, by Chebyshev's inequality, we have

$$P\left(\left|\operatorname{tr}(P_s^*\Omega) - \tilde{\epsilon}^{'}P_s^*\tilde{\epsilon}\right| > \delta\zeta_n^*\right) \leq \delta^{-2}\zeta_n^{*-2}\operatorname{Var}(\tilde{\epsilon}^{'}P_s^*\tilde{\epsilon}) = o(1), \tag{E.21}$$

by Assumption 9 and the fact that $\operatorname{Var}(\tilde{\epsilon}^{'}P_s^*\tilde{\epsilon}) = O(n)$, which is ensured by Lemma A.3 and Assumption 4. For the term, $\zeta_n^{*-1}\left|\operatorname{tr}\left((\widetilde{P}_s - P_s^*)\widehat{\Omega}\right)\right|$, we have

$$\begin{aligned}\zeta_n^{*-1}\left|\operatorname{tr}\left((\widetilde{P}_s - P_s^*)\widehat{\Omega}\right)\right| &\leq \zeta_n^{*-1}\sup_s\left(\gamma_{\max}(\widetilde{P}_s - P_s^*)\gamma_{\max}(\widehat{\Omega})\operatorname{rank}(\widetilde{P}_s - P_s^*)\right) \\ &\leq 2k\zeta_n^{*-1}\sup_s\gamma_{\max}(\widetilde{P}_s - P_s^*)\gamma_{\max}(\widehat{\Omega}) = o_p(1),\end{aligned} \tag{E.22}$$

by Assumption 9. Similarly, by Assumption 9 and Lemma A.3, we have

$$\zeta_n^{*-1}\left|\text{tr}\left(P_s^*(\widehat{\Omega}-\Omega)\right)\right| \leq \zeta_n^{*-1}\sup_s\left(\gamma_{\max}(\widehat{\Omega}-\Omega)\gamma_{\max}(P_s^*)\text{rank}(P_s^*)\right)$$
$$\leq k\zeta_n^{*-1}\left(\gamma_{\max}(\widehat{\Omega})+\gamma_{\max}(\Omega)\right)\sup_s\gamma_{\max}(P_s^*) = o_p(1). \tag{E.23}$$

Finally, for the third term, using the bounds of Rayleigh quotient, we obtain $\zeta_n^{*-1}\left|\tilde{\epsilon}'(\widetilde{P}_s-P_s^*)\tilde{\epsilon}\right| \leq \zeta_n^{*-1}\|\tilde{\epsilon}\|^2\sup_s\gamma_{\max}(\widetilde{P}_s-P_s^*) = o_p(1)$ by Assumption 9 and Lemma A.4. Thus, $\zeta_n^{*-1}\left|\text{tr}(\widetilde{P}_s\widehat{\Omega})-\tilde{\epsilon}'\widetilde{P}_s\tilde{\epsilon}\right| = o_p(1)$.

Next, we consider $\zeta_n^{*-1}|L_s-R_s^*|$. From the proof Theorem 1, it will be enough to show that

$$(a)\ \left\|(\widetilde{P}_s-P_s^*)\mu\right\|^2/\zeta_n^* = o_p(1),\quad (b)\ \left\|(\widetilde{P}_s-P_s^*)\tilde{\epsilon}\right\|^2/\zeta_n^* = o_p(1),$$
$$(c)\ \left|\|P_s^*\tilde{\epsilon}\|^2-\text{tr}(P_s^*\Omega P_s^{*'})\right|/\zeta_n^* = o_p(1),\quad (d)\ \left|\mu'(\widetilde{P}_s-P_s^*)'(\widetilde{P}_s-P_s^*)\tilde{\epsilon}\right|/\zeta_n^* = o_p(1),$$
$$(e)\ \left|\mu'(\widetilde{P}_s-P_s^*)'P_s^*\tilde{\epsilon}\right|/\zeta_n^* = o_p(1),\quad (f)\ \left|\mu'(\widetilde{P}_s-P_s^*)'H_s^*\mu\right|/\zeta_n^* = o_p(1),$$
$$(g)\ \left|\tilde{\epsilon}'(\widetilde{P}_s-P_s^*)'P_s^*\tilde{\epsilon}\right|/\zeta_n^* = o_p(1),\quad (h)\ \left|\tilde{\epsilon}'(\widetilde{P}_s-P_s^*)'H_s^*\mu\right|/\zeta_n^* = o_p(1),$$
$$(j)\ \left|\tilde{\epsilon}'P_s^{*'}H_s^*\mu\right|/\zeta_n^* = o_p(1).$$

We can prove the above results by using similar approaches to those used in (E.19)-E.23. Thus, we have shown that $\widehat{C}_s = R_s^*+q_n$, where $q_n/\zeta_n^* = o_p(1)$ uniformly for $s\in\mathcal{S}$. Note that Assumption 9 ensures that $\zeta_n^*\to\infty$. Therefore, uniformly for $s\in\mathcal{S}$, we have $\widehat{C}_s = R_s^*+q_n$ with $R_s^*\geq\zeta_n^*\to\infty$, which completes the proof.

E.3 Proof of Theorem 3

Similar to the proof of Theorem 1, we need to verify the following:

$$\sup_z R^*(z)^{-1}\left|\tilde{\epsilon}'\widetilde{H}(z)\mu\right| = o_p(1), \tag{E.24}$$

$$\sup_z R^*(z)^{-1}\left|\text{tr}(\widetilde{P}(z)\widehat{\Omega})-\tilde{\epsilon}'\widetilde{P}(z)\tilde{\epsilon}\right| = o_p(1), \tag{E.25}$$

$$\sup_z R^*(z)^{-1}\left|\sum_{s=1}^S z_s\frac{\partial\widehat{\alpha}_s}{\partial y'}\widehat{\Omega}\frac{\partial\widetilde{P}_s}{\partial\widehat{\alpha}_s}y\right| = o_p(1), \tag{E.26}$$

$$\sup_z R^*(z)^{-1}\left|\sum_{s=1}^S z_s\frac{\partial\widehat{\tau}_s}{\partial y'}\widehat{\Omega}\frac{\partial\widetilde{P}_s}{\partial\widehat{\tau}_s}y\right| = o_p(1), \tag{E.27}$$

$$\sup_z R^*(z)^{-1}|L(z)-R^*(z)| = o_p(1), \tag{E.28}$$

where $z \in \mathcal{N}$, $\sup_z$ denote supremum over $\mathcal{N}$ and $\widetilde{H}(z) = I_n - \widetilde{P}(z)$. For (E.24), similar to (E.3), we need to prove

$$\sup_z R^*(z)^{-1} \left| \tilde{\epsilon}' H^*(z)\mu \right| = o_p(1), \tag{E.29}$$

and

$$\sup_z R^*(z)^{-1} \left| \tilde{\epsilon}' \left(P^*(z) - \widetilde{P}(z) \right) \mu \right| = o_p(1), \tag{E.30}$$

where $H^*(z) = I_n - P^*(z)$. Note that $H^*(z) = \sum_{s=1}^S z_s H_s^*$ because $\sum_{s=1}^S z_s = 1$. Denote z_s^o as an $S \times 1$ vector of zeros except the sth element which is one. For (E.29), by a similar logic to the first term in (E.3), we obtain

$$\begin{aligned}
&P\left(\sup_z R^*(z)^{-1} \left| \tilde{\epsilon}' H^*(z)\mu \right| > \eta \right) \le P\left(\tilde{\zeta}_n^{-1} \sup_z \sum_{s=1}^S z_s \left| \tilde{\epsilon}' H_s^* \mu \right| > \eta \right) \\
&= P\left(\tilde{\zeta}_n^{-1} \max_{1 \le s \le S} \left| \tilde{\epsilon}' H_s^* \mu \right| > \eta \right) = P\left(\left(\tilde{\zeta}_n^{-1} \left| \tilde{\epsilon}' H^*(z_1^o)\mu \right| > \eta \right) \cup \cdots \cup \left(\tilde{\zeta}_n^{-1} \left| \tilde{\epsilon}' H^*(z_S^o)\mu \right| > \eta \right) \right) \\
&\le \sum_{s=1}^S P\left(\tilde{\zeta}_n^{-1} \left| \tilde{\epsilon}' H^*(z_s^o)\mu \right| > \eta \right) \le \sum_{s=1}^S \mathbb{E}\left(\frac{\left(\tilde{\epsilon}' H^*(z_s^o)\mu \right)^{2G}}{\eta^{2G} \tilde{\zeta}_n^{2G}} \right) \\
&\le c_3 \eta^{-2G} \tilde{\zeta}_n^{-2G} \sum_{s=1}^S \| H^*(z_s^o)\mu \|^{2G} \le c_3 \eta^{-2G} \tilde{\zeta}_n^{-2G} \sum_{s=1}^S \left(R^*(z_s^o) \right)^{2G} = o(1),
\end{aligned}$$

where c_3 is a constant, the second inequality follows from the Boole's inequality and the last equality from Assumption 10. For (E.30), using the Cauchy-Schwarz inequality and the bounds of Rayleigh quotient, we obtain

$$\begin{aligned}
&\sup_z R^*(z)^{-1} \left| \tilde{\epsilon}' \left(P^*(z) - \widetilde{P}(z) \right) \mu \right| \le \tilde{\zeta}_n^{-1} \|\tilde{\epsilon}\| \gamma_{\max} \left(\sum_{s=1}^S z_s (P_s^* - \widetilde{P}_s) \right) \|\mu\| \\
&\le \tilde{\zeta}_n^{-1} \|\tilde{\epsilon}\| \sum_{s=1}^S z_s \gamma_{\max} \left(P_s^* - \widetilde{P}_s \right) \|\mu\| \le \|\tilde{\epsilon}\| \tilde{\zeta}_n^{-1} \sup_s \gamma_{\max} \left(P_s^* - \widetilde{P}_s \right) \|\mu\| = o_p(1),
\end{aligned}$$

by Assumptions 6 and 10, and the fact that $\|\tilde{\epsilon}\| = O_p(n^{1/2})$, which is ensured by Lemma A.4. For (E.25), similar to E.4, we have to prove the following results:

$$\sup_z R^*(z)^{-1} \left| \operatorname{tr}(P^*(z)\Omega) - \tilde{\epsilon}' P^*(z)\tilde{\epsilon} \right| = o_p(1), \tag{E.31}$$

$$\sup_z R^*(z)^{-1} \left| \operatorname{tr}\left(P^*(z)(\widehat{\Omega} - \Omega) \right) \right| = o_p(1), \tag{E.32}$$

$$\sup_z R^*(z)^{-1} \left| \operatorname{tr}\left(\left(\widetilde{P}(z) - P^*(z) \right) \widehat{\Omega} \right) \right| = o_p(1), \tag{E.33}$$

$$\sup_z R^*(z)^{-1} \left| \tilde{\epsilon}' \left(\widetilde{P}(z) - P^*(z) \right) \tilde{\epsilon} \right| = o_p(1). \tag{E.34}$$

Define $D(z_s^o) = P^*(z_s^o)e^{-\tau M}e^{-\alpha W}$. For (E.31), using a similar approach to the one used in the second term in (E.4), we obtain

$$\begin{aligned} P\left(\sup_z R^*(z)^{-1} \left| \operatorname{tr}(P^*(z)\Omega) - \tilde{\epsilon}' P^*(z)\tilde{\epsilon} \right| > \eta \right) &\leq \sum_{s=1}^{S} P\left(\tilde{\zeta}_n^{-1} \left| \operatorname{tr}(P^*(z_s^o)\Omega) - \tilde{\epsilon}' P^*(z_s^o)\tilde{\epsilon} \right| > \eta \right) \\ \leq \sum_{s=1}^{S} \mathbb{E}\left(\frac{\left(\operatorname{tr}(P^*(z_s^o)\Omega) - \tilde{\epsilon}' P^*(z_s^o)\tilde{\epsilon} \right)^{2G}}{\eta^{2G}\tilde{\zeta}_n^{2G}} \right) &\leq c_4 \eta^{-2G} \tilde{\zeta}_n^{-2G} \sum_{s=1}^{S} \left(\operatorname{tr}\left(D(z_s^o) D(z_s^o)' \right) \right)^G \\ \leq c_4' \eta^{-2G} \tilde{\zeta}_n^{-2G} \sum_{s=1}^{S} \left(R^*(z_s^o) \right)^{-G} = o_p(1), & \end{aligned}$$

for some constants c_4 and c_4'. Similarly, the results (E.32)-(E.34) can be obtained by similar approaches used in the other terms in (E.4). Moreover, (E.26) and (E.27) can be easily obtained from the second part of Assumption 10.

To prove (E.28), similar to (E.7), we need to show the following results:

$$\sup_z R^*(z)^{-1} \left\| \left(\widetilde{P}(z) - P^*(z) \right) \mu \right\|^2 = o_p(1), \tag{E.35}$$

$$\sup_z R^*(z)^{-1} \left\| \left(\widetilde{P}(z) - P^*(z) \right) \epsilon \right\|^2 = o_p(1), \tag{E.36}$$

$$\sup_z R^*(z)^{-1} \left| \|P^*(z)\tilde{\epsilon}\|^2 - \operatorname{tr}\left(P^*(z) \Omega P^{*'}(z) \right) \right| = o_p(1), \tag{E.37}$$

$$\sup_z R^*(z)^{-1} \left| \mu' \left(\widetilde{P}(z) - P^*(z) \right)' \left(\widetilde{P}(z) - P^*(z) \right) \tilde{\epsilon} \right| = o_p(1), \tag{E.38}$$

$$\sup_z R^*(z)^{-1} \left| \mu' \left(\widetilde{P}(z) - P^*(z) \right)' P^*(z) \tilde{\epsilon} \right| = o_p(1), \tag{E.39}$$

$$\sup_z R^*(z)^{-1} \left| \mu' \left(\widetilde{P}(z) - P^*(z) \right)' H^*(z) \mu \right| = o_p(1), \tag{E.40}$$

$$\sup_z R^*(z)^{-1} \left| \tilde{\epsilon}' \left(\widetilde{P}(z) - P^*(z) \right)' P^*(z) \tilde{\epsilon} \right| = o_p(1), \tag{E.41}$$

$$\sup_z R^*(z)^{-1} \left| \tilde{\epsilon}' \left(\widetilde{P}(z) - P^*(z) \right)' H^*(z) \mu \right| = o_p(1), \tag{E.42}$$

$$\sup_z R^*(z)^{-1} \left| \mu' P^{*'}(z) H^*(z) \mu \right| = o_p(1). \tag{E.43}$$

For (E.35), using the bounds of Rayleigh quotient, we obtain

$$\begin{aligned}
&\sup_z R^*(z)^{-1} \left\| \left(\widetilde{P}(z) - P^*(z) \right) \mu \right\|^2 \leq \tilde{\zeta}_n^{-1} \gamma_{\max} \left(\left(\widetilde{P}(z) - P^*(z) \right)' \left(\widetilde{P}(z) - P^*(z) \right) \right) \|\mu\|^2 \\
&\leq \tilde{\zeta}_n^{-1} \gamma_{\max}^2 \left(\left(\widetilde{P}(z) - P^*(z) \right) \right) \|\mu\|^2 \leq \tilde{\zeta}_n^{-1} \left(\sum_{s=1}^{S} z_s \gamma_{\max}(\widetilde{P}_s - P_s^*) \right)^2 \|\mu\|^2 \\
&\leq \tilde{\zeta}_n^{-1} \sup_s \gamma_{\max}^2 (\widetilde{P}_s - P_s^*) \|\mu\|^2 = o_p(1),
\end{aligned}$$

by Assumptions 6 and 10. The result in (E.36) can be obtained similarly.

To prove the result in (E.37), first define $D(z_s^o) = P^*(z_s^o) e^{-\alpha W} e^{-\tau M}$. Then, using a similar approach to the one used in the first term of (E.4), we obtain

$$
\begin{aligned}
& P\left(\sup_z R^*(z)^{-1}\left|\|P^*(z)\tilde{\epsilon}\|^2-\operatorname{tr}\left(P^*(z)\Omega P^{*'}(z)\right)\right|>\eta\right) \\
& \leq P\left(\tilde{\zeta}_n^{-1}\sup_z\sum_{t=1}^S\sum_{s=1}^S z_t z_s\left|\tilde{\epsilon}^{'}P_t^{*'}P_s^*\tilde{\epsilon}-\operatorname{tr}(P_s^*\Omega P_t^{*'})\right|>\eta\right) \\
& \leq P\left(\tilde{\zeta}_n^{-1}\max_{1\leq t\leq S}\max_{1\leq s\leq S}\left|\tilde{\epsilon}^{'}P_t^{*'}P_s^*\tilde{\epsilon}-\operatorname{tr}(P_s^*\Omega P_t^{*'})\right|>\eta\right) \\
& \leq P\left(\tilde{\zeta}_n^{-1}\left|\tilde{\epsilon}^{'}P^{*'}(z_1^o)P^*(z_1^o)\tilde{\epsilon}-\operatorname{tr}(P^*(z_1^o)\Omega P^{*'}(z_1^o))\right|>\eta\right) \\
& \quad \cup P\left(\tilde{\zeta}_n^{-1}\left|\tilde{\epsilon}^{'}P^{*'}(z_1^o)P^*(z_2^o)\tilde{\epsilon}-\operatorname{tr}(P^*(z_2^o)\Omega P^{*'}(z_1^o))\right|>\eta\right)\cdots \\
& \quad \cup P\left(\tilde{\zeta}_n^{-1}\left|\tilde{\epsilon}^{'}P^{*'}(z_1^o)P^*(z_S^o)\tilde{\epsilon}-\operatorname{tr}(P^*(z_S^o)\Omega P^{*'}(z_1^o))\right|>\eta\right) \\
& \quad \cup P\left(\tilde{\zeta}_n^{-1}\left|\tilde{\epsilon}^{'}P^{*'}(z_2^o)P^*(z_1^o)\tilde{\epsilon}-\operatorname{tr}(P^*(z_1^o)\Omega P^{*'}(z_2^o))\right|>\eta\right)\cdots \\
& \quad \cup P\left(\tilde{\zeta}_n^{-1}\left|\tilde{\epsilon}^{'}P^{*'}(z_2^o)P^*(z_S^o)\tilde{\epsilon}-\operatorname{tr}(P^*(z_S^o)\Omega P^{*'}(z_2^o))\right|>\eta\right)\cdots \\
& \quad \cup P\left(\tilde{\zeta}_n^{-1}\left|\tilde{\epsilon}^{'}P^{*'}(z_S^o)P^*(z_S^o)\tilde{\epsilon}-\operatorname{tr}(P^*(z_S^o)\Omega P^{*'}(z_S^o))\right|>\eta\right) \\
& \leq \sum_{t=1}^S\sum_{s=1}^S P\left(\tilde{\zeta}_n^{-1}\left|\tilde{\epsilon}^{'}P^{*'}(z_t^o)P^*(z_s^o)\tilde{\epsilon}-\operatorname{tr}(P^*(z_s^o)\Omega P^{*'}(z_t^o))\right|>\eta\right) \\
& \leq \sum_{t=1}^S\sum_{s=1}^S\frac{\mathbb{E}\left(\tilde{\epsilon}^{'}P^{*'}(z_t^o)P^*(z_s^o)\tilde{\epsilon}-\operatorname{tr}(P^*(z_s^o)\Omega P^{*'}(z_t^o))\right)^{2G}}{\tilde{\zeta}_n^{2G}\eta^{2G}} \\
& \leq c_5\tilde{\zeta}^{-2G}\eta^{-2G}\sum_{t=1}^S\sum_{s=1}^S\left(\operatorname{tr}\left(D(z_s^o)D^{'}(z_s^o)\right)\right)^G \\
& \leq c_5^{'}\tilde{\zeta}^{-2G}\eta^{-2G}S\sum_{s=1}^S(R_s^*)^G \\
& = o(1),
\end{aligned}
$$

where c_5 and $c_5^{'}$ are some constant terms, the fourth inequality follows from the Boole's inequality and the last equality from Assumption 10. Finally, the results in (E.38)-(E.43) can be obtained similarly to those in (E.14)-(E.18). This completes the proof.

References

Acemoglu, Daron et al. (2012). "The Network Origins of Aggregate Fluctuations". In: *Econometrica* 80.5, pp. 1977–2016.

Anselin, Luc (1986). "Non-nested tests on the weight structure in spatial autoregressive models: Some Monte Carlo results". In: *Journal of Regional Science* 26.2, pp. 267–284.

— (1988). *Spatial econometrics: Methods and Models*. New York: Springer.

Anselin, Luc et al. (1996). "Simple diagnostic tests for spatial dependence". In: *Regional Science and Urban Economics* 26.1, pp. 77–104.

Arbia, Giuseppe et al. (2020). "Testing Impact Measures in Spatial Autoregressive Models". In: *International Regional Science Review* 43.1-2, pp. 40–75.

Bailey, Natalia et al. (2016). "A Two-Stage Approach to Spatio-Temporal Analysis with Strong and Weak Cross-Sectional Dependence". In: *Journal of Applied Econometrics* 31.1, pp. 249–280.

Baltagi, Badi H. and Dong Li (2001). "Double Length Artificial Regressions For Testing Spatial Dependence". In: *Econometric Reviews* 20.1, pp. 31–40.

Baltagi, Badi H. and Long Liu (2016). "Random Effects, Fixed Effects and Hausman's Test for the Generalized Mixed Regressive Spatial Autoregressive Panel Data Model". In: *Econometric Reviews* 35.4, pp. 638–658.

Baltagi, Badi H. and Zhenlin Yang (2013). "Heteroskedasticity and non-normality robust LM tests for spatial dependence". In: *Regional Science and Urban Economics* 43.5, pp. 725 –739.

Baltagi, Badi H. et al. (2007). "Estimating models of complex FDI: Are there third-country effects?" In: *Journal of Econometrics* 140.1. Analysis of spatially dependent data, pp. 260 –281.

— (2008). "Estimating regional trade agreement effects on FDI in an interdependent world". In: *Journal of Econometrics* 145.1. The use of econometrics in informing public policy makers, pp. 194 –208.

Baltagi, Badi H. et al. (2016). "Firm-Level Productivity Spillovers in China's Chemical Industry: A Spatial Hausman-Taylor Approach". In: *Journal of Applied Econometrics* 31.1, pp. 214–248.

— (2017). "Determinants of firm-level domestic sales and exports with spillovers: Evidence from China". In: *Journal of Econometrics* 199.2. The Creative Mind in Econometrics: Studies in Celebration of Robert Basmann's 90th Year on Causation, Identification and Structural Equation Estimation, pp. 184 –201.

Behrens, Kristian et al. (2012). "'Dual' Gravity: Using Spatial Econometrics to Control for Multilateral Resistance". In: *Journal of Applied Econometrics* 27.5, pp. 773–794.

Born, Benjamin and Jörg Breitung (2011). "Simple regression-based tests for spatial dependence". In: *The Econometrics Journal* 14.2, pp. 330–342.

Bramoullé, Yann et al. (2009). "Identification of peer effects through social networks". In: *Journal of Econometrics* 150.1, pp. 41 –55.

Burridge, Peter (2012). "Improving the J Test in the SARAR Model by Likelihood-based Estimation". In: *Spatial Economic Analysis* 7.1, pp. 75–107.

Burridge, Peter and Bernard Fingleton (2010). “Bootstrap Inference in Spatial Econometrics: the J-test”. In: *Spatial Economic Analysis* 5.1, pp. 93–119.

Calvo-Armengol, Antoni et al. (2009). “Peer Effects and Social Networks in Education”. In: *The Review of Economic Studies* 76.4, pp. 1239–1267.

Case, Anne C. (1991). “Spatial Patterns in Household Demand”. In: *Econometrica* 59.4, pp. 953–965.

Chiu, Tom YM et al. (1996). “The matrix-logarithmic covariance model”. In: *Journal of the American Statistical Association* 91.433, pp. 198–210.

Corrado, Luisa and Bernard Fingleton (2012). “Where is the Economics in Spatial Econometrics?” In: *Journal of Regional Science* 52.2, pp. 210–239.

Debarsy, Nicolas and James P. LeSage (2020). “Bayesian Model Averaging for Spatial Autoregressive Models Based on Convex Combinations of Different Types of Connectivity Matrices”. In: *Journal of Business & Economic Statistics* 0.0, pp. 1–12.

Debarsy, Nicolas et al. (2015). “Large sample properties of the matrix exponential spatial specification with an application to FDI”. In: *Journal of Econometrics* 188.1, pp. 1–21.

Desmet, Klaus and Esteban Rossi-Hansberg (2009). “Spatial growth and industry age”. In: *Journal of Economic Theory* 144.6. Dynamic General Equilibrium, pp. 2477 –2502.

Doğan, Osman et al. (2018). “Simple tests for social interaction models with network structures”. In: *Spatial Economic Analysis* 13.2, pp. 212–246.

Ertur, Cem and Wilfried Koch (2007). “Growth, technological interdependence and spatial externalities: theory and evidence”. In: *Journal of applied econometrics* 22.6, pp. 1033–1062.

— (2011). “A contribution to the theory and empirics of Schumpeterian growth with worldwide interactions”. In: *Journal of Economic Growth* 16.3, pp. 215–255.

Feng, Yang et al. (2020). “On the sparsity of Mallows model averaging estimator”. In: *Economics Letters* 187, p. 108916.

Gao, Yan et al. (2019). “Frequentist model averaging for threshold models”. In: *Annals of the Institute of Statistical Mathematics* 71.2, pp. 275–306.

Han, Xiaoyi and Lung fei Lee (2013). “Model selection using J-test for the spatial autoregressive model vs. the matrix exponential spatial model”. In: *Regional Science and Urban Economics* 43.2, pp. 250 –271.

Han, Xiaoyi et al. (2017). “Estimation and model selection of higher-order spatial autoregressive model: An efficient Bayesian approach”. In: *Regional Science and Urban Economics* 63, pp. 97 –120.

Hansen, Bruce E. (2007). “Least squares model averaging”. In: *Econometrica* 75.4, pp. 1175–1189.

— (2014). “Model averaging, asymptotic risk, and regressor groups”. In: *Quantitative Economics* 5.3, pp. 495–530.

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

Jin, Fei and Lung fei Lee (2013). "Cox-type tests for competing spatial autoregressive models with spatial autoregressive disturbances". In: *Regional Science and Urban Economics* 43.4, pp. 590 –616.

Kelejian, Harry H. (2008). "A spatial J-test for model specification against a single or a set of non-nested alternatives". In: *Letters in Spatial and Resource Sciences* 1, pp. 3 –11.

Kelejian, Harry H. and Gianfranco Piras (2011). "An extension of Kelejian's J-test for non-nested spatial models". In: *Regional Science and Urban Economics* 41.3, pp. 281 –292.

— (2014). "Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes". In: *Regional Science and Urban Economics* 46.0, pp. 140–149.

Kelejian, Harry H. and Ingmar R. Prucha (1998). "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances". In: *The Journal of Real Estate Finance and Economics* 17, 99–121.

— (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* 157.1, pp. 53 –67.

Kelejian, Harry H. et al. (2013). "Spatial spillovers in the development of institutions". In: *Journal of Development Economics* 101.0, pp. 297 –315.

König, Michael D. et al. (2019). "R&D Networks: Theory, Empirics, and Policy Implications". In: *The Review of Economics and Statistics* 101.3, pp. 476–491.

Lee, Lung-Fei (2003). "Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances". In: *Econometric Reviews* 22.4, pp. 307–335.

— (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 (2012). "Convergence: A Spatial Dynamic Panel Data Approach". In: *Global Journal of Economics* 1.1, p. 36.

Lee, Lung-Fei et al. (2020). "Who is the Key Player? A Network Analysis of Juvenile Delinquency". In: *Journal of Business & Economic Statistics* 0.0, pp. 1–9.

Leeb, Hannes and Benedikt M. Pötscher (2005). "Model Selection and Inference: Facts and Fiction". In: *Econometric Theory* 21.1, pp. 21–59.

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

LeSage, James P. (2014). "Spatial econometric panel data model specification: A Bayesian approach". In: *Spatial Statistics* 9. Revealing Intricacies in Spatial and Spatio-Temporal Data: Papers from the Spatial Statistics 2013 Conference, pp. 122 –145.

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

Li, Ker-Chau (1987). "Asymptotic optimality for Cp, CL, cross-validation and generalized cross-validation: discrete index set". In: *The Annals of Statistics*, pp. 958–975.

Liang, Hua et al. (2011). "Optimal weight choice for frequentist model average estimators". In: *Journal of the American Statistical Association* 106.495, pp. 1053–1066.

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

Liu, Qingfeng and Ryo Okui (2013). "Heteroscedasticity-robust C_p model averaging". In: *The Econometrics Journal* 16.3, pp. 463–472.

Mallows, C.L. (1973). "Some comments on C_p". In: *Technometrics* 15, pp. 661–675.

Patacchini, Eleonora and Yves Zenou (2016). "Social Networks And Parental Behavior In The Intergenerational Transmission Of Religion". In: *Quantitative Economics* 7.3, pp. 969–995.

Pesaran, M. Hashem and Cynthia Fan Yang (2020). "Econometric analysis of production networks with dominant units". In: *Journal of Econometrics* 219.2, pp. 507 –541.

Stein, Charles M (1981). "Estimation of the mean of a multivariate normal distribution". In: *The annals of Statistics*, pp. 1135–1151.

Taşpınar, Süleyman et al. (2018). "GMM inference in spatial autoregressive models". In: *Econometric Reviews* 37.9, pp. 931–954.

Wan, Alan TK et al. (2010). "Least squares model averaging by Mallows criterion". In: *Journal of Econometrics* 156.2, pp. 277–283.

Whittle, Peter (1960). "Bounds for the moments of linear and quadratic forms in independent variables". In: *Theory of Probability & Its Applications* 5.3, pp. 302–305.

Zhang, Xinyu (2021). "Optimal model averaging based on generalized method of moments". In: *Statistica Sinica* 31.4. DOI: `http://dx.doi.org/10.5705/ss.202019.0230`.

Zhang, Xinyu and Jihai Yu (2018). "Spatial weights matrix selection and model averaging for spatial autoregressive models". In: *Journal of Econometrics* 203.1, pp. 1–18.

Zhang, Xinyu et al. (2014). "Model averaging and weight choice in linear mixed-effects models". In: *Biometrika* 101.1, pp. 205–218.

Web Appendix for Model Selection and Model Averaging for Matrix Exponential Spatial Models

Ye Yang* Osman Doğan† Süleyman Taşpınar‡

February 11, 2022

In this appendix, we provide additional simulation results on the finite sample performance of our suggested MS and MA procedures. The simulation setting for these results is described in the main text. For the normal distribution case, Tables 1 and 2 report the simulation results when the true spatial weights matrices are W_2 and W_3, respectively. For the non-normal distribution case, Tables 3, 4 and 5 report the simulation results when the true spatial weights matrices are W_2, W_3 and W_4, respectively. The remaining tables, Tables 6–8, include the simulation results for the heteroskedastic case.

*School of Accounting, Capital University of Economics and Business, Beijing, China, email: yang.ye@cueb.edu.cn
†Department of Economics, University of Illinois, Illinois, U.S.A., email: odogan@illinois.edu
‡Department of Economics, Queens College, The City University of New York, New York, U.S.A., email: staspinar@qc.cuny.edu

Table 1: True W is W_2 under normal disturbance

			W_1	W_2	W_3	W_4	MS	MA
α=0.2	n=169	RMSE of α	0.112	0.049	0.062	0.087	0.054	
τ=0.2		RMSE of τ	0.145	0.093	0.124	0.201	0.101	
β_1=2		RMSE of β_1	0.095	0.075	0.079	0.089	0.075	
β_2=1		RMSE of β_2	0.083	0.078	0.082	0.083	0.079	
		Loss	13.817	3.430	10.814	15.625	4.568	4.330
		MS accuracy	0.057	0.858	0.071	0.014		
		MA weights	0.103	0.729	0.103	0.065		
α=0.2	n=400	RMSE of α	0.112	0.035	0.048	0.058	0.035	
τ=0.2		RMSE of τ	0.125	0.061	0.087	0.157	0.062	
β_1=2		RMSE of β_1	0.089	0.054	0.056	0.056	0.054	
β_2=1		RMSE of β_2	0.051	0.050	0.062	0.067	0.051	
		Loss	25.017	3.459	21.223	30.787	3.985	4.330
		MS accuracy	0.006	0.971	0.023	0.000		
		MA weights	0.062	0.836	0.076	0.026		
			W_1	W_2	W_3	W_4	MS	MA
α=−0.2	n=169	RMSE of α	0.102	0.046	0.075	0.114	0.057	
τ=−0.2		RMSE of τ	0.137	0.095	0.137	0.234	0.115	
β_1=2		RMSE of β_1	0.089	0.079	0.102	0.118	0.080	
β_2=1		RMSE of β_2	0.071	0.071	0.081	0.097	0.071	
		Loss	13.727	3.306	13.127	18.737	4.845	4.500
		MS accuracy	0.064	0.867	0.054	0.015		
		MA weights	0.105	0.752	0.085	0.058		
α=−0.2	n=400	RMSE of α	0.106	0.031	0.043	0.081	0.033	
τ=−0.2		RMSE of τ	0.121	0.061	0.083	0.185	0.063	
β_1=2		RMSE of β_1	0.082	0.052	0.060	0.055	0.052	
β_2=1		RMSE of β_2	0.067	0.051	0.052	0.054	0.051	
		Loss	25.880	3.259	22.714	34.457	3.839	4.158
		MS accuracy	0.013	0.974	0.009	0.004		
		MA weights	0.063	0.850	0.050	0.037		
			W_1	W_2	W_3	W_4	MS	MA
α=1.2	n=169	RMSE of α	1.001	0.060	3.474	4.606	0.070	
τ=1.2		RMSE of τ	0.844	0.104	3.499	4.665	0.176	
β_1=2		RMSE of β_1	1.142	0.102	0.780	0.867	0.116	
β_2=1		RMSE of β_2	0.575	0.061	0.550	0.464	0.081	
		Loss	1592.040	23.011	913.465	926.259	57.168	65.376
		MS accuracy	0.001	0.947	0.027	0.025		
		MA weights	0.049	0.838	0.065	0.048		
α=1.2	n=400	RMSE of α	1.013	0.038	2.421	4.403	0.039	
τ=1.2		RMSE of τ	0.848	0.063	2.607	4.458	0.071	
β_1=2		RMSE of β_1	1.163	0.065	0.721	0.817	0.066	
β_2=1		RMSE of β_2	0.586	0.044	0.416	0.507	0.047	
		Loss	3972.496	25.020	2092.509	2277.550	33.755	69.030
		MS accuracy	0.000	0.994	0.004	0.002		
		MA weights	0.027	0.894	0.050	0.029		
			W_1	W_2	W_3	W_4	MS	MA
α=−1.2	n=169	RMSE of α	1.056	0.067	3.006	11.491	0.434	
τ=−1.2		RMSE of τ	0.740	0.106	3.004	11.627	0.613	
β_1=2		RMSE of β_1	1.153	0.111	0.567	1.308	0.148	
β_2=1		RMSE of β_2	0.602	0.066	0.321	0.702	0.100	
		Loss	1420.104	25.696	669.128	1262.617	65.657	74.989
		MS accuracy	0.000	0.946	0.025	0.029		
		MA weights	0.052	0.854	0.021	0.073		
α=−1.2	n=400	RMSE of α	1.027	0.037	1.689	13.408	0.044	
τ=−1.2		RMSE of τ	0.807	0.065	1.828	13.819	0.177	
β_1=2		RMSE of β_1	1.139	0.061	0.380	1.425	0.064	
β_2=1		RMSE of β_2	0.565	0.037	0.211	0.718	0.049	
		Loss	3316.116	18.657	1341.870	3219.301	31.253	66.618
		MS accuracy	0.000	0.992	0.002	0.006		
		MA weights	0.033	0.905	0.015	0.047		

Table 2: True W is W_3 under normal error terms

			W_1	W_2	W_3	W_4	MS	MA
α=0.2	n=169	RMSE of α	0.147	0.095	0.064	0.083	0.070	
τ=0.2		RMSE of τ	0.100	0.146	0.127	0.178	0.142	
β_1=2		RMSE of β_1	0.079	0.078	0.077	0.086	0.077	
β_2=1		RMSE of β_2	0.086	0.086	0.083	0.085	0.084	
		Loss	10.769	7.680	3.424	8.226	4.970	4.407
		MS accuracy	0.026	0.108	0.703	0.163		
		MA weights	0.066	0.119	0.624	0.192		
α=0.2	n=400	RMSE of α	0.151	0.099	0.043	0.055	0.048	
τ=0.2		RMSE of τ	0.073	0.119	0.084	0.118	0.092	
β_1=2		RMSE of β_1	0.053	0.055	0.051	0.051	0.051	
β_2=1		RMSE of β_2	0.049	0.049	0.048	0.051	0.048	
		Loss	17.751	12.381	2.957	12.469	4.391	4.037
		MS accuracy	0.008	0.062	0.855	0.075		
		MA weights	0.044	0.092	0.745	0.118		
			W_1	W_2	W_3	W_4	MS	MA
α=−0.2	n=169	RMSE of α	0.157	0.117	0.073	0.102	0.085	
τ=−0.2		RMSE of τ	0.093	0.125	0.142	0.196	0.151	
β_1=2		RMSE of β_1	0.082	0.080	0.078	0.083	0.080	
β_2=1		RMSE of β_2	0.074	0.074	0.071	0.078	0.073	
		Loss	9.518	7.472	3.309	6.726	4.817	4.206
		MS accuracy	0.060	0.137	0.697	0.106		
		MA weights	0.098	0.151	0.614	0.137		
α=−0.2	n=400	RMSE of α	0.150	0.100	0.044	0.077	0.049	
τ=−0.2		RMSE of τ	0.072	0.116	0.083	0.137	0.088	
β_1=2		RMSE of β_1	0.058	0.051	0.049	0.051	0.049	
β_2=1		RMSE of β_2	0.051	0.049	0.048	0.049	0.048	
		Loss	19.005	13.434	2.975	13.022	3.704	3.712
		MS accuracy	0.008	0.033	0.931	0.028		
		MA weights	0.047	0.081	0.785	0.087		
			W_1	W_2	W_3	W_4	MS	MA
α=1.2	n=169	RMSE of α	0.974	0.611	0.062	2.534	0.167	
τ=1.2		RMSE of τ	0.338	0.735	0.127	2.643	0.259	
β_1=2		RMSE of β_1	0.322	0.208	0.088	0.465	0.105	
β_2=1		RMSE of β_2	0.303	0.146	0.065	0.239	0.088	
		Loss	472.162	257.049	10.187	466.014	42.516	33.931
		MS accuracy	0.003	0.045	0.906	0.046		
		MA weights	0.034	0.026	0.853	0.088		
α=1.2	n=400	RMSE of α	0.963	0.597	0.042	3.096	0.102	
τ=1.2		RMSE of τ	0.356	0.731	0.083	3.189	0.149	
β_1=2		RMSE of β_1	0.374	0.133	0.056	0.473	0.064	
β_2=1		RMSE of β_2	0.198	0.089	0.046	0.296	0.048	
		Loss	1035.937	549.931	10.224	982.778	28.952	36.242
		MS accuracy	0.000	0.018	0.973	0.009		
		MA weights	0.029	0.019	0.894	0.058		
			W_1	W_2	W_3	W_4	MS	MA
α=−1.2	n=169	RMSE of α	0.981	0.582	0.064	0.939	0.248	
τ=−1.2		RMSE of τ	0.416	0.794	0.129	0.775	0.265	
β_1=2		RMSE of β_1	0.273	0.360	0.089	0.261	0.109	
β_2=1		RMSE of β_2	0.143	0.245	0.061	0.162	0.074	
		Loss	439.287	243.193	10.680	336.684	41.285	37.177
		MS accuracy	0.017	0.063	0.887	0.033		
		MA weights	0.062	0.040	0.830	0.069		
α=−1.2	n=400	RMSE of α	1.047	0.674	0.041	0.543	0.108	
τ=−1.2		RMSE of τ	0.320	0.694	0.083	0.305	0.109	
β_1=2		RMSE of β_1	0.212	0.304	0.054	0.158	0.058	
β_2=1		RMSE of β_2	0.069	0.259	0.040	0.141	0.044	
		Loss	995.274	557.724	9.241	443.054	19.920	28.424
		MS accuracy	0.000	0.010	0.982	0.008		
		MA weights	0.060	0.023	0.902	0.016		

Table 3: True W is W_2 under non-normal (stadanrdized χ^2_3) error terms

			W_1	W_2	W_3	W_4	MS	MA
α=0.2	n=169	RMSE of α	0.112	0.049	0.064	0.086	0.053	
τ=0.2		RMSE of τ	0.141	0.091	0.126	0.202	0.097	
β_1=2		RMSE of β_1	0.094	0.077	0.082	0.093	0.077	
β_2=1		RMSE of β_2	0.087	0.080	0.084	0.086	0.081	
		Loss	13.879	3.499	11.027	15.756	4.718	4.427
		MS accuracy	0.042	0.860	0.079	0.019		
		MA weights	0.090	0.735	0.110	0.066		
α=0.2	n=400	RMSE of α	0.104	0.030	0.047	0.064	0.031	
τ=0.2		RMSE of τ	0.119	0.056	0.081	0.165	0.057	
β_1=2		RMSE of β_1	0.070	0.052	0.054	0.056	0.052	
β_2=1		RMSE of β_2	0.065	0.049	0.052	0.052	0.049	
		Loss	26.065	3.124	24.085	37.562	3.563	3.992
		MS accuracy	0.008	0.979	0.013	0.000		
		MA weights	0.059	0.846	0.072	0.024		
			W_1	W_2	W_3	W_4	MS	MA
α=−0.2	n=169	RMSE of α	0.115	0.050	0.066	0.156	0.062	
τ=−0.2		RMSE of τ	0.148	0.096	0.125	0.236	0.109	
β_1=2		RMSE of β_1	0.108	0.078	0.088	0.086	0.079	
β_2=1		RMSE of β_2	0.086	0.077	0.080	0.088	0.078	
		Loss	13.551	3.305	12.155	17.897	4.640	4.357
		MS accuracy	0.049	0.884	0.044	0.023		
		MA weights	0.100	0.761	0.079	0.060		
α=−0.2	n=400	RMSE of α	0.111	0.034	0.046	0.074	0.037	
τ=−0.2		RMSE of τ	0.125	0.061	0.083	0.190	0.069	
β_1=2		RMSE of β_1	0.073	0.051	0.053	0.057	0.051	
β_2=1		RMSE of β_2	0.056	0.052	0.054	0.057	0.052	
		Loss	25.896	3.382	21.811	29.757	4.071	4.396
		MS accuracy	0.004	0.970	0.016	0.010		
		MA weights	0.058	0.841	0.046	0.055		
			W_1	W_2	W_3	W_4	MS	MA
α=1.2	n=169	RMSE of α	1.012	0.057	3.325	5.062	0.079	
τ=1.2		RMSE of τ	0.745	0.098	3.294	5.049	0.146	
β_1=2		RMSE of β_1	1.090	0.091	0.585	0.766	0.105	
β_2=1		RMSE of β_2	0.586	0.066	0.459	0.437	0.085	
		Loss	1528.231	19.706	795.227	931.889	59.420	61.765
		MS accuracy	0.001	0.943	0.032	0.024		
		MA weights	0.049	0.856	0.057	0.038		
α=1.2	n=400	RMSE of α	1.014	0.039	2.382	4.217	0.039	
τ=1.2		RMSE of τ	0.853	0.062	2.582	4.257	0.075	
β_1=2		RMSE of β_1	1.164	0.066	0.595	0.801	0.066	
β_2=1		RMSE of β_2	0.575	0.041	0.425	0.436	0.043	
		Loss	3966.367	24.997	1799.558	2469.166	36.853	67.524
		MS accuracy	0.000	0.992	0.006	0.002		
		MA weights	0.031	0.896	0.051	0.022		
			W_1	W_2	W_3	W_4	MS	MA
α=−1.2	n=169	RMSE of α	0.965	0.053	2.305	11.718	0.138	
τ=−1.2		RMSE of τ	0.693	0.092	2.372	12.386	0.353	
β_1=2		RMSE of β_1	1.130	0.095	0.487	1.368	0.127	
β_2=1		RMSE of β_2	0.558	0.058	0.387	0.749	0.115	
		Loss	1639.635	20.355	752.891	1463.676	58.905	72.958
		MS accuracy	0.004	0.952	0.013	0.031		
		MA weights	0.050	0.844	0.028	0.078		
α=−1.2	n=400	RMSE of α	1.007	0.039	1.145	12.929	0.044	
τ=−1.2		RMSE of τ	0.842	0.062	1.279	13.141	0.098	
β_1=2		RMSE of β_1	1.155	0.067	0.634	1.440	0.068	
β_2=1		RMSE of β_2	0.578	0.045	0.366	0.760	0.046	
		Loss	3889.033	25.745	1829.504	3612.798	28.537	74.325
		MS accuracy	0.000	0.999	0.000	0.001		
		MA weights	0.029	0.906	0.033	0.033		

Table 4: True W is W_3 under non-normal (stadanrdized χ_3^2) error terms

			W_1	W_2	W_3	W_4	MS	MA
α=0.2	n=169	RMSE of α	0.147	0.093	0.061	0.079	0.068	
τ=0.2		RMSE of τ	0.106	0.148	0.124	0.172	0.140	
β_1=2		RMSE of β_1	0.080	0.077	0.076	0.082	0.077	
β_2=1		RMSE of β_2	0.082	0.082	0.081	0.082	0.081	
		Loss	10.641	7.394	3.164	8.026	4.660	4.081
		MS accuracy	0.031	0.112	0.713	0.144		
		MA weights	0.068	0.118	0.645	0.169		
α=0.2	n=400	RMSE of α	0.153	0.109	0.047	0.061	0.050	
τ=0.2		RMSE of τ	0.073	0.109	0.085	0.119	0.089	
β_1=2		RMSE of β_1	0.055	0.058	0.051	0.053	0.052	
β_2=1		RMSE of β_2	0.052	0.052	0.051	0.052	0.051	
		Loss	17.742	12.984	3.230	15.365	4.229	4.149
		MS accuracy	0.007	0.038	0.911	0.044		
		MA weights	0.041	0.071	0.789	0.098		
			W_1	W_2	W_3	W_4	MS	MA
α=−0.2	n=169	RMSE of α	0.159	0.109	0.067	0.131	0.084	
τ=−0.2		RMSE of τ	0.093	0.134	0.127	0.216	0.144	
β_1=2		RMSE of β_1	0.086	0.078	0.076	0.078	0.078	
β_2=1		RMSE of β_2	0.087	0.083	0.079	0.084	0.080	
		Loss	10.715	8.230	3.142	9.075	4.392	3.979
		MS accuracy	0.033	0.084	0.817	0.066		
		MA weights	0.073	0.119	0.707	0.101		
α=−0.2	n=400	RMSE of α	0.154	0.103	0.048	0.071	0.054	
τ=−0.2		RMSE of τ	0.070	0.115	0.082	0.128	0.092	
β_1=2		RMSE of β_1	0.051	0.053	0.049	0.052	0.049	
β_2=1		RMSE of β_2	0.058	0.058	0.056	0.058	0.056	
		Loss	18.229	12.923	3.359	10.780	4.454	4.307
		MS accuracy	0.010	0.056	0.882	0.052		
		MA weights	0.061	0.101	0.744	0.094		
			W_1	W_2	W_3	W_4	MS	MA
α=1.2	n=169	RMSE of α	0.990	0.622	0.062	2.598	0.104	
τ=1.2		RMSE of τ	0.336	0.796	0.128	2.709	0.216	
β_1=2		RMSE of β_1	0.332	0.269	0.082	0.391	0.101	
β_2=1		RMSE of β_2	0.216	0.138	0.069	0.259	0.077	
		Loss	485.307	273.603	9.173	414.074	29.691	28.458
		MS accuracy	0.001	0.011	0.945	0.043		
		MA weights	0.039	0.023	0.865	0.074		
α=1.2	n=400	RMSE of α	0.974	0.595	0.040	4.298	0.099	
τ=1.2		RMSE of τ	0.382	0.766	0.083	4.406	0.151	
β_1=2		RMSE of β_1	0.329	0.195	0.057	0.567	0.061	
β_2=1		RMSE of β_2	0.118	0.148	0.043	0.330	0.045	
		Loss	1097.609	575.632	10.141	1217.466	30.486	38.014
		MS accuracy	0.000	0.017	0.973	0.010		
		MA weights	0.032	0.018	0.898	0.053		
			W_1	W_2	W_3	W_4	MS	MA
α=−1.2	n=169	RMSE of α	0.926	0.679	0.061	0.535	0.211	
τ=−1.2		RMSE of τ	0.421	0.914	0.122	0.551	0.283	
β_1=2		RMSE of β_1	0.357	0.268	0.087	0.164	0.096	
β_2=1		RMSE of β_2	0.139	0.201	0.058	0.151	0.078	
		Loss	610.846	290.374	12.499	248.331	48.779	44.771
		MS accuracy	0.004	0.062	0.875	0.059		
		MA weights	0.076	0.044	0.815	0.065		
α=−1.2	n=400	RMSE of α	0.979	0.614	0.041	0.735	0.104	
τ=−1.2		RMSE of τ	0.374	0.730	0.084	0.413	0.130	
β_1=2		RMSE of β_1	0.367	0.126	0.057	0.147	0.058	
β_2=1		RMSE of β_2	0.185	0.089	0.045	0.096	0.046	
		Loss	905.125	509.739	8.556	587.272	19.063	26.593
		MS accuracy	0.000	0.019	0.979	0.002		
		MA weights	0.044	0.029	0.901	0.027		

Table 5: True W is W_4 under non-normal (stadanrdized χ^2_3) error terms

			W_1	W_2	W_3	W_4	MS	MA
α=0.2	n=169	RMSE of α	0.173	0.146	0.100	0.076	0.085	
τ=0.2		RMSE of τ	0.178	0.208	0.162	0.174	0.184	
β_1=2		RMSE of β_1	0.074	0.073	0.073	0.074	0.074	
β_2=1		RMSE of β_2	0.082	0.082	0.082	0.081	0.081	
		Loss	7.228	6.372	5.224	2.710	3.984	3.464
		MS accuracy	0.062	0.117	0.115	0.706		
		MA weights	0.095	0.120	0.115	0.669		
α=0.2	n=400	RMSE of α	0.182	0.166	0.109	0.063	0.072	
τ=0.2		RMSE of τ	0.155	0.173	0.122	0.118	0.129	
β_1=2		RMSE of β_1	0.059	10.227	0.055	0.051	0.051	
β_2=1		RMSE of β_2	0.050	0.050	0.050	0.050	0.050	
		Loss	10.227	9.617	7.920	2.763	4.089	3.616
		MS accuracy	0.024	0.055	0.106	0.815		
		MA weights	0.068	0.079	0.113	0.739		

			W_1	W_2	W_3	W_4	MS	MA
$\alpha=-0.2$	n=169	RMSE of α	0.187	0.174	0.158	0.120	0.142	
$\tau=-0.2$		RMSE of τ	0.116	0.134	0.151	0.196	0.178	
β_1=2		RMSE of β_1	0.077	0.075	0.077	0.076	0.076	
β_2=1		RMSE of β_2	0.080	0.078	0.077	0.075	0.078	
		Loss	6.544	6.364	5.651	3.496	5.105	4.379
		MS accuracy	0.102	0.135	0.216	0.547		
		MA weights	0.129	0.143	0.228	0.501		
$\alpha=-0.2$	n=400	RMSE of α	0.173	0.144	0.112	0.069	0.080	
$\tau=-0.2$		RMSE of τ	0.108	0.137	0.134	0.126	0.129	
β_1=2		RMSE of β_1	0.050	0.050	0.049	0.049	0.049	
β_2=1		RMSE of β_2	0.057	0.058	0.057	0.055	0.056	
		Loss	12.729	10.860	7.814	3.515	4.820	4.426
		MS accuracy	0.040	0.049	0.152	0.759		
		MA weights	0.081	0.080	0.167	0.672		

			W_1	W_2	W_3	W_4	MS	MA
α=1.2	n=169	RMSE of α	1.076	0.931	0.365	0.063	0.080	
τ=1.2		RMSE of τ	1.242	1.401	0.884	0.177	0.198	
β_1=2		RMSE of β_1	0.168	0.177	0.157	0.079	0.080	
β_2=1		RMSE of β_2	0.177	0.154	0.195	0.074	0.074	
		Loss	122.422	103.227	194.649	3.694	4.137	6.478
		MS accuracy	0.000	0.003	0.001	0.996		
		MA weights	0.034	0.041	0.032	0.892		
α=1.2	n=400	RMSE of α	1.083	0.981	0.274	0.042	0.042	
τ=1.2		RMSE of τ	1.286	1.393	0.806	0.110	0.110	
β_1=2		RMSE of β_1	0.163	0.152	0.180	0.053	0.053	
β_2=1		RMSE of β_2	0.144	0.125	0.130	0.044	0.044	
		Loss	282.270	255.846	616.782	3.634	3.634	5.714
		MS accuracy	0.000	0.000	0.000	1.000		
		MA weights	0.014	0.027	0.013	0.946		

			W_1	W_2	W_3	W_4	MS	MA
$\alpha=-1.2$	n=169	RMSE of α	1.065	0.820	1.088	0.130	0.317	
$\tau=-1.2$		RMSE of τ	0.274	0.485	0.285	0.203	0.239	
β_1=2		RMSE of β_1	0.173	0.266	0.183	0.078	0.088	
β_2=1		RMSE of β_2	0.176	0.108	0.196	0.077	0.084	
		Loss	457.505	307.583	439.141	22.171	55.116	45.632
		MS accuracy	0.004	0.072	0.025	0.899		
		MA weights	0.034	0.039	0.080	0.846		
$\alpha=-1.2$	n=400	RMSE of α	1.113	0.932	1.233	0.091	0.242	
$\tau=-1.2$		RMSE of τ	0.219	0.394	0.201	0.136	0.153	
β_1=2		RMSE of β_1	0.265	0.080	0.318	0.058	0.065	
β_2=1		RMSE of β_2	0.109	0.074	0.152	0.048	0.050	
		Loss	583.407	465.628	659.137	13.725	38.044	37.814
		MS accuracy	0.004	0.041	0.005	0.950		
		MA weights	0.034	0.035	0.059	0.872		

Table 6: True W is W_2 under heteroskedastic error terms

			W_1	W_2	W_3	W_4	MS	MA
α=0.2	n=169	RMSE of α	0.113	0.050	0.066	0.091	0.056	
τ=0.2		RMSE of τ	0.148	0.097	0.134	0.221	0.105	
β_1=2		RMSE of β_1	0.114	0.102	0.109	0.118	0.103	
β_2=1		RMSE of β_2	0.085	0.078	0.083	0.086	0.079	
		Loss	14.706	4.496	12.064	16.794	5.908	5.487
		MS accuracy	0.056	0.836	0.090	0.018		
		MA weights	0.097	0.720	0.118	0.066		
α=0.2	n=400	RMSE of α	0.102	0.030	0.049	0.065	0.033	
τ=0.2		RMSE of τ	0.118	0.059	0.087	0.173	0.061	
β_1=2		RMSE of β_1	0.083	0.068	0.070	0.071	0.068	
β_2=1		RMSE of β_2	0.062	0.047	0.050	0.049	0.047	
		Loss	26.686	3.952	24.991	38.339	4.469	4.857
		MS accuracy	0.012	0.976	0.012	0.000		
		MA weights	0.061	0.844	0.072	0.022		

			W_1	W_2	W_3	W_4	MS	MA
α=−0.2	n=169	RMSE of α	0.118	0.050	0.069	0.161	0.065	
τ=−0.2		RMSE of τ	0.144	0.092	0.123	0.246	0.107	
β_1=2		RMSE of β_1	0.127	0.104	0.115	0.111	0.104	
β_2=1		RMSE of β_2	0.088	0.083	0.086	0.095	0.083	
		Loss	14.601	4.409	13.317	19.131	6.147	5.531
		MS accuracy	0.074	0.842	0.061	0.023		
		MA weights	0.106	0.751	0.081	0.062		
α=−0.2	n=400	RMSE of α	0.112	0.034	0.050	0.071	0.036	
τ=−0.2		RMSE of τ	0.127	0.062	0.088	0.190	0.067	
β_1=2		RMSE of β_1	0.084	0.065	0.067	0.070	0.065	
β_2=1		RMSE of β_2	0.059	0.056	0.058	0.062	0.056	
		Loss	26.888	4.325	22.924	30.529	4.868	5.324
		MS accuracy	0.006	0.975	0.013	0.006		
		MA weights	0.061	0.840	0.051	0.049		

			W_1	W_2	W_3	W_4	MS	MA
α=1.2	n=169	RMSE of α	1.025	0.065	2.327	5.928	0.075	
τ=1.2		RMSE of τ	0.750	0.104	2.392	5.953	0.168	
β_1=2		RMSE of β_1	1.067	0.120	0.594	0.984	0.130	
β_2=1		RMSE of β_2	0.629	0.065	0.396	0.581	0.079	
		Loss	1796.327	35.367	765.477	1242.728	74.591	80.532
		MS accuracy	0.001	0.949	0.021	0.029		
		MA weights	0.046	0.850	0.059	0.045		
α=1.2	n=400	RMSE of α	1.011	0.037	3.237	4.501	0.037	
τ=1.2		RMSE of τ	0.850	0.064	3.422	4.580	0.066	
β_1=2		RMSE of β_1	1.153	0.076	0.644	0.737	0.076	
β_2=1		RMSE of β_2	0.581	0.039	0.340	0.385	0.039	
		Loss	4049.869	27.181	1893.256	2528.386	32.350	74.028
		MS accuracy	0.000	0.996	0.004	0.000		
		MA weights	0.030	0.900	0.047	0.022		

			W_1	W_2	W_3	W_4	MS	MA
α=−1.2	n=169	RMSE of α	0.950	0.058	1.518	11.544	0.138	
τ=−1.2		RMSE of τ	0.685	0.096	1.609	11.766	0.342	
β_1=2		RMSE of β_1	1.135	0.116	0.813	1.329	0.139	
β_2=1		RMSE of β_2	0.534	0.065	0.310	0.682	0.108	
		Loss	1327.822	22.200	871.228	1367.843	58.519	72.662
		MS accuracy	0.001	0.951	0.020	0.028		
		MA weights	0.043	0.846	0.036	0.076		
α=−1.2	n=400	RMSE of α	1.002	0.042	1.271	12.435	0.042	
τ=−1.2		RMSE of τ	0.801	0.065	1.436	12.727	0.071	
β_1=2		RMSE of β_1	1.147	0.079	0.638	1.398	0.079	
β_2=1		RMSE of β_2	0.574	0.044	0.386	0.740	0.044	
		Loss	3864.587	30.816	1859.131	3502.113	34.836	86.543
		MS accuracy	0.000	0.997	0.003	0.000		
		MA weights	0.030	0.898	0.037	0.035		

Table 7: True W is W_3 under heteroskedastic error terms

			W_1	W_2	W_3	W_4	MS	MA
α=0.2	n=169	RMSE of α	0.147	0.094	0.066	0.085	0.074	
τ=0.2		RMSE of τ	0.104	0.146	0.130	0.180	0.143	
β_1=2		RMSE of β_1	0.102	0.100	0.100	0.106	0.100	
β_2=1		RMSE of β_2	0.080	0.081	0.079	0.085	0.080	
		Loss	11.471	8.379	4.191	9.007	5.871	5.212
		MS accuracy	0.047	0.137	0.677	0.139		
		MA weights	0.078	0.142	0.608	0.171		
α=0.2	n=400	RMSE of α	0.153	0.109	0.046	0.062	0.051	
τ=0.2		RMSE of τ	0.071	0.107	0.085	0.122	0.091	
β_1=2		RMSE of β_1	0.072	0.075	0.069	0.070	0.069	
β_2=1		RMSE of β_2	0.050	0.050	0.049	0.051	0.049	
		Loss	18.411	13.691	4.004	16.206	5.210	4.947
		MS accuracy	0.009	0.051	0.895	0.045		
		MA weights	0.042	0.080	0.783	0.095		
			W_1	W_2	W_3	W_4	MS	MA
α=−0.2	n=169	RMSE of α	0.158	0.109	0.067	0.135	0.082	
τ=−0.2		RMSE of τ	0.099	0.134	0.129	0.222	0.147	
β_1=2		RMSE of β_1	0.112	0.106	0.106	0.107	0.106	
β_2=1		RMSE of β_2	0.086	0.086	0.083	0.088	0.084	
		Loss	11.857	9.300	4.252	10.273	5.837	5.302
		MS accuracy	0.059	0.110	0.775	0.056		
		MA weights	0.102	0.120	0.684	0.094		
α=−0.2	n=400	RMSE of α	0.153	0.103	0.048	0.068	0.053	
τ=−0.2		RMSE of τ	0.073	0.117	0.088	0.130	0.096	
β_1=2		RMSE of β_1	0.066	0.068	0.064	0.067	0.064	
β_2=1		RMSE of β_2	0.058	0.058	0.056	0.058	0.056	
		Loss	18.995	13.671	4.128	11.360	5.178	4.985
		MS accuracy	0.010	0.057	0.881	0.052		
		MA weights	0.063	0.100	0.747	0.090		
			W_1	W_2	W_3	W_4	MS	MA
α=1.2	n=169	RMSE of α	0.968	0.725	0.066	4.101	0.142	
τ=1.2		RMSE of τ	0.370	0.966	0.132	4.180	0.231	
β_1=2		RMSE of β_1	0.420	0.235	0.108	0.605	0.114	
β_2=1		RMSE of β_2	0.202	0.196	0.065	0.347	0.075	
		Loss	569.807	265.604	15.509	547.940	37.734	36.543
		MS accuracy	0.001	0.042	0.933	0.024		
		MA weights	0.034	0.037	0.853	0.076		
α=1.2	n=400	RMSE of α	0.976	0.801	0.044	2.359	0.093	
τ=1.2		RMSE of τ	0.357	0.880	0.084	2.467	0.123	
β_1=2		RMSE of β_1	0.166	0.351	0.070	0.373	0.073	
β_2=1		RMSE of β_2	0.185	0.120	0.041	0.208	0.042	
		Loss	1109.869	615.068	13.879	1169.966	27.535	38.434
		MS accuracy	0.000	0.013	0.982	0.005		
		MA weights	0.032	0.016	0.903	0.049		
			W_1	W_2	W_3	W_4	MS	MA
α=−1.2	n=169	RMSE of α	0.979	1.314	0.062	0.873	0.202	
τ=−1.2		RMSE of τ	0.431	1.460	0.128	0.671	0.230	
β_1=2		RMSE of β_1	0.382	0.284	0.109	0.262	0.122	
β_2=1		RMSE of β_2	0.281	0.185	0.073	0.192	0.090	
		Loss	411.606	268.487	10.454	295.791	30.358	29.320
		MS accuracy	0.009	0.030	0.927	0.034		
		MA weights	0.065	0.027	0.846	0.061		
α=−1.2	n=400	RMSE of α	0.982	0.614	0.042	0.752	0.127	
τ=−1.2		RMSE of τ	0.370	0.728	0.081	0.438	0.137	
β_1=2		RMSE of β_1	0.372	0.141	0.072	0.164	0.073	
β_2=1		RMSE of β_2	0.187	0.089	0.044	0.092	0.047	
		Loss	909.669	513.615	10.855	597.735	26.466	34.005
		MS accuracy	0.000	0.028	0.970	0.002		
		MA weights	0.042	0.033	0.896	0.030		

Table 8: True W is W_4 under heteroskedastic error terms

			W_1	W_2	W_3	W_4	MS	MA
α=0.2	n=169	RMSE of α	0.174	0.149	0.102	0.084	0.092	
τ=0.2		RMSE of τ	0.177	0.204	0.157	0.183	0.193	
β_1=2		RMSE of β_1	0.100	0.099	0.099	0.099	0.100	
β_2=1		RMSE of β_2	0.083	0.083	0.082	0.081	0.082	
		Loss	8.081	7.214	6.169	3.613	4.806	4.415
		MS accuracy	0.057	0.081	0.129	0.733		
		MA weights	0.094	0.108	0.124	0.673		
α=0.2	n=400	RMSE of α	0.183	0.168	0.113	0.067	0.076	
τ=0.2		RMSE of τ	0.154	0.171	0.118	0.119	0.126	
β_1=2		RMSE of β_1	0.073	10.805	0.070	0.066	0.067	
β_2=1		RMSE of β_2	0.047	0.047	0.047	0.047	0.047	
		Loss	10.805	10.206	8.548	3.481	4.501	4.228
		MS accuracy	0.033	0.041	0.067	0.859		
		MA weights	0.067	0.077	0.090	0.765		
			W_1	W_2	W_3	W_4	MS	MA
α=−0.2	n=169	RMSE of α	0.187	0.172	0.156	0.124	0.141	
τ=−0.2		RMSE of τ	0.123	0.141	0.155	0.202	0.188	
β_1=2		RMSE of β_1	0.102	0.101	0.102	0.101	0.101	
β_2=1		RMSE of β_2	0.080	0.081	0.082	0.079	0.082	
		Loss	7.463	7.205	6.639	4.490	6.066	5.261
		MS accuracy	0.113	0.097	0.243	0.547		
		MA weights	0.141	0.112	0.235	0.512		
α=−0.2	n=400	RMSE of α	0.175	0.144	0.112	0.067	0.079	
τ=−0.2		RMSE of τ	0.110	0.141	0.141	0.119	0.130	
β_1=2		RMSE of β_1	0.069	0.069	0.069	0.069	0.069	
β_2=1		RMSE of β_2	0.058	0.059	0.058	0.058	0.058	
		Loss	13.816	11.920	9.016	4.522	5.995	5.486
		MS accuracy	0.039	0.053	0.171	0.737		
		MA weights	0.080	0.075	0.176	0.668		
			W_1	W_2	W_3	W_4	MS	MA
α=1.2	n=169	RMSE of α	1.059	0.919	0.315	0.060	0.086	
τ=1.2		RMSE of τ	1.267	1.407	0.873	0.176	0.208	
β_1=2		RMSE of β_1	0.140	0.131	0.146	0.098	0.099	
β_2=1		RMSE of β_2	0.164	0.157	0.146	0.072	0.072	
		Loss	125.120	104.991	244.548	4.366	5.014	6.744
		MS accuracy	0.000	0.006	0.000	0.994		
		MA weights	0.031	0.042	0.018	0.910		
α=1.2	n=400	RMSE of α	1.121	1.048	0.400	0.045	0.045	
τ=1.2		RMSE of τ	1.225	1.302	0.792	0.110	0.110	
β_1=2		RMSE of β_1	0.293	0.278	0.232	0.066	0.066	
β_2=1		RMSE of β_2	0.082	0.091	0.072	0.040	0.040	
		Loss	238.350	224.707	482.115	4.583	4.583	6.920
		MS accuracy	0.000	0.000	0.000	1.000		
		MA weights	0.017	0.026	0.014	0.942		
			W_1	W_2	W_3	W_4	MS	MA
α=−1.2	n=169	RMSE of α	1.123	0.988	1.261	0.152	0.396	
τ=−1.2		RMSE of τ	0.225	0.356	0.277	0.217	0.232	
β_1=2		RMSE of β_1	0.268	0.149	0.386	0.110	0.126	
β_2=1		RMSE of β_2	0.098	0.108	0.153	0.072	0.074	
		Loss	232.002	199.282	272.159	15.398	39.293	32.021
		MS accuracy	0.035	0.070	0.014	0.881		
		MA weights	0.064	0.046	0.065	0.826		
α=−1.2	n=400	RMSE of α	1.113	0.931	1.226	0.100	0.219	
τ=−1.2		RMSE of τ	0.217	0.394	0.186	0.139	0.154	
β_1=2		RMSE of β_1	0.270	0.093	0.318	0.073	0.077	
β_2=1		RMSE of β_2	0.110	0.074	0.147	0.046	0.047	
		Loss	584.700	466.925	654.524	17.011	35.280	37.972
		MS accuracy	0.003	0.034	0.001	0.962		
		MA weights	0.032	0.032	0.051	0.885		