Fast Estimation of Matrix Exponential Spatial Models

Ye Yang* Osman Doğan† Süleyman Taşpınar‡

December 17, 2021

Abstract

The matrix exponential spatial specification (MESS) is an alternative to the spatial autoregressive-type (SAR-type) specifications with several attractive properties. The spatial dependence in the MESS-type models is formulated through a matrix exponential term, and the estimation of these models may require the computation of the matrix exponential terms many times in an estimation procedure. In the literature, it is well documented that the computation of the matrix exponential terms can pose challenges in terms of reliability, stability, accuracy, and efficiency. We propose a matrix-vector products approach based on the truncation of Taylor series expansion of the matrix exponential terms for the fast estimation of MESS-type models. We show how to efficiently implement this approach for a first-order MESS model, and provide extensive simulation evidence for its computational advantage over the default method utilized by a popular statistical software.

JEL-Classification: C13, C21, C31.
Keywords: Matrix Exponential, MESS, QML, GMM, Bayesian, Inference, Impact Measures.

*Department of Economics, The Graduate Center, The City University of New York, New York, U.S.A., email: yyang1@gradcenter.cuny.edu

†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 account for potential (weak) cross-sectional dependence among observations located on some relevant space. The Cliff-Ord-type spatial models (i.e., the spatial autoregressive (SAR)-type) impose an autoregressive specification for the variable of interest using a weights matrix which quantifies the strength of a measure of nearness between spatial units. (Anselin, 1988; Cliff and Ord, 1969, 1973; Whittle, 1954). The spatial autoregressive specification ensures that the spatial dependence among spatial units decay at a geometric rate. The likelihood based estimation of these models requires calculation of matrix determinant terms (the Jacobian terms) in each pass of the numerical optimization scheme, and for large matrices that are not sparse this can create computational challenges. See for example LeSage and Pace (2009, Chapter 4) for a variety of approximation methods suggested in the literature to alleviate this computational problem.

An alternative to SAR-type models is the matrix exponential spatial specification (MESS) suggested by LeSage and Pace (2007). In the MESS-type models, the spatial dependence is formulated with a matrix exponential term of type, $e^{\alpha A} = \sum_{i=0}^{\infty}(\alpha A)^i/i!$, where α is a scalar spatial parameter, A is the $n \times n$ spatial weights matrix and n is the number of spatial units. Therefore, the MESS-type models impose an exponential rate of decay for the cross-sectional dependence and have several features that make them more convenient for estimation (Chiu et al., 1996; Leonard and Hsu, 1992). The likelihood based estimation is greatly simplified because the likelihood function does not involve any matrix determinant terms. Since matrix exponential terms are always invertible, there is no need to impose restrictions on the parameter space of the spatial parameters, i.e., the MESS-type models always have reduced forms. Moreover, when the model involves heteroskedasticity of an unknown form in the error terms, the maximum likelihood estimator (MLE) remains consistent provided that the weights matrices are commutative (Debarsy et al., 2015).[1]

Despite the aforementioned advantages of the MESS-type models, the likelihood and generalized method of moments (GMM) based estimations of these models require the computation of matrix exponential terms by the optimization solvers in each iteration, which can be computationally costly. The literature has suggested several alternative ways to compute $e^{\alpha A}$ such as Taylor series approximation, Padé approximation, ordinary differential equation methods, polynomial methods, matrix decomposition methods, splitting methods and Krylov space methods. Moler and Van Loan (1978, 2003) assess the effectiveness of nineteen methods according to the following attributes: (i) generality, (ii) reliability, (iii) stability, (iv) accuracy, (v) efficiency, (vi) storage requirements, (vii) ease of use, and (viii) simplicity. Although Moler and Van Loan (1978, 2003) state that "none (of the methods in their paper) are completely satisfactory", they claim that a scaling and squaring method with either the rational Padé or Taylor approximants can be the most effective one to compute the matrix exponential terms.[2] Popular software such as Python, R, MATLAB and Mathematica

[1]See Debarsy et al. (2015) for the formal results on on the maximum likelihood (ML) and generalized method of moments (GMM) estimation of the MESS models.

[2]On the scaling and squaring method with either the rational Padé or Taylor approximants, see also Higham

have modules, packages and functions to compute matrix exponential of a given matrix. For example, MATLAB (function `expm`), Mathematica (function `MatrixExp`) and Python (function `scipy.linalg.expm`) use a scaling and squaring method combined with a Padé approximation for the computation of matrix exponential terms (Al-Mohy and Higham, 2010; Higham, 2005).

As pointed out by Moler and Van Loan (1978, 2003), all methods suggested in the literature for computing the matrix exponential terms are "dubious" in the sense that a sole method may not be entirely reliable for all applications.[3] In other words, a method that is effective for a particular application may not be reliable for an another application. In the context of MESS-type models, the scaling and squaring method combined with the Padé approximation as implemented in MATLAB through its `expm` function can be highly costly in terms of computation time. For example, a typical Monte Carlo simulation designed for a first-order MESS model with 1000 resampling and a couple of different parameter combinations can take days or even weeks.

In this paper, we propose a matrix-vector products method based on the truncation of Taylor series expansion of the matrix exponential terms (the first method in Moler and Van Loan (1978, 2003)) for fast estimation of MESS-type models. Our analysis on the estimation of MESS-type models indicate that the estimation requires the computation of a matrix exponential term as a vector, rather than the matrix exponential term in isolation. For example, the estimation of a first-order MESS model requires the computation of terms such as $e^{\alpha A}e^{\tau B}v$ and $e^{\tau B}X$ where v is an $n \times 1$ vector, X is an $n \times k$ matrix, and α and τ are the spatial parameters. The matrix-vector products method provides approximations to $e^{\alpha A}e^{\tau B}v$ and $e^{\tau B}X$ in terms of matrix-vector products rather than providing approximations for $e^{\alpha A}$ and $e^{\tau B}$. In this paper, we show how this approach can be implemented for the MESS-type models. Using the suggested approach in the context of a first-order MESS model, we provide extensive simulation evidence on computational time gains for three estimation methods (the QML, GMM and Bayesian methods).

In the literature, it is well known that the matrix-vector products approach can reduce the computational burden substantially. For example, Moler and Van Loan (2003) write "One of the most significant changes in numerical linear algebra in the past 25 years is the rise of iterative methods for sparse matrix problems, in which only matrix vector products are needed." In particular, LeSage and Pace (2009) extensively use sparse matrix-vector operations for the likelihood and Bayesian estimation of spatial models to reduce the computational burden. The Krylov space methods (the twentieth method in Moler and Van Loan (2003)) suggested for the computation of matrix exponential terms also depend on the matrix-vector products approach. See, for example, Saad (1992), Gallopoulos and Saad (1992), Hochbruck and Lubich (1997), Sidje (1998), and related references. When A is sparse, in the first step of this approach, $e^{\alpha A}v$ is approximated by an element of Krylov subspace $K_m = \text{span}\{v, (\alpha A)v, \ldots, (\alpha A)^m v\}$, where m, the dimension of the Krylov subspace, is

(2005) and Bader et al. (2019). Sidje (1998) provide an extensive package named `ExpoKIT` (both the Fortran and MATLAB versions are available) to compute matrix exponential terms with the Krylov subspace method using the Arnoldi process approximation.

[3]The computation of matrix exponential terms may be necessary for many applications from different fields. For example, MATLAB uses its `expm` function in its Control Toolbox, System Identification Toolbox, Neural Net Toolbox, Mu-Analysis and Synthesis Toolbox, Model Predictive Control toolbox, and Simulink.

small compare to n. The operations in the first step of this method involves only matrix-vector products (see Sidje (1998) for the implementation of this method).

The remainder of this paper is organized as follows. Section 2 presents the model under consideration and lays out briefly the details on the QML, GMM, and Bayesian MCMC estimation. This section also shows how the impact measures and their dispersion measures can be estimated. Section 3 provides the details on the computation of the matrix exponential terms using the matrix-vector products approach. We then show how the matrix-vector products approach can be applied to the QML, GMM and Bayesian estimation methods. Section 4 presents the setting for our Monte Carlo study and the simulation results. Section 5 illustrates the computational time advantage of the matrix-vector products method using a large dataset from the spatial econometric literature. We conclude in Section 6. Some simulation results are relegated to an appendix.

2 Model and Estimation Approaches

2.1 Model Specification

We consider the following first order matrix exponential spatial model (for short MESS$(1,1)$)

$$e^{\alpha_0 W} y = X\beta_0 + u, \quad e^{\tau_0 M} u = \epsilon, \tag{2.1}$$

where $Y = (y_1, \ldots, y_n)^{'}$ is the $n \times 1$ vector of observations on the dependent variable, X is the $n \times k$ matrix of non-stochastic exogenous variables with the associated parameter vector β_0, W and M are the $n \times n$ spatial weights matrices of known constants with zero diagonal elements. The scalar parameters α_0 and τ_0 are called the spatial parameters. We call $U = (u_1, \ldots, u_n)^{'}$ as the $n \times 1$ vector of regression disturbance terms and $\epsilon = (\epsilon_1, \ldots, \epsilon_n)^{'}$ as the $n \times 1$ vector of disturbances (or innovations).

The matrix exponential terms $e^{\alpha W}$ and $e^{\tau M}$ in (2.1) are defined as $e^{\alpha W} = \sum_{i=0}^{\infty}(\alpha W)^i/i!$ and $e^{\tau M} = \sum_{i=0}^{\infty}(\tau M)^i/i!$, and are always invertable with the inverses $e^{-\alpha W}$ and $e^{-\tau M}$ (Chiu et al., 1996). Thus, the reduced form of the model always exists and is given by $y = e^{-\alpha W}X\beta_0 + e^{-\alpha W}e^{-\tau M}\epsilon$. On the other hand, in a SAR-type model, we need to restrict the parameter space of spatial parameters so that it has a reduced form. On the parameter space of spatial parameters in the SAR-type models, among others, see Elhorst (2014), Kelejian and Prucha (2010), and LeSage and Pace (2009). In the SAR-type counterpart of the MESS(1,1) model, the matrix exponential terms $e^{\alpha_0 W}y$ and $e^{\tau_0 M}u$ in (2.1) are respectively replaced by $(I_n - \lambda_0 W)y$ and $(I_n - \rho_0 M)u$, where I_n is the $n \times n$ identity matrix, and λ_0 and ρ_0 are scalar spatial parameters. Under the assumption that $\|\lambda_0 W\| < 1$ for some matrix norm $\|\cdot\|$, we have $(I_n - \lambda_0 W)^{-1} = \sum_{i=0}^{\infty}(\lambda_0 W)^i$ (Horn and Johnson, 2012). Thus, the SAR model imposes a geometric decay pattern of spatial dependence among spatial units, while the MESS$(1,1)$ model exhibits an exponential decay.[4]

[4]On the properties of MESS and SAR type models, see LeSage and Pace (2007, 2009) and Debarsy et al. (2015).

2.2 Maximum Likelihood Approach

Under the assumption that ϵ_i's are i.i.d normal with mean zero and variance σ_0^2, the log-likelihood function of the model can be expressed as

$$\ln L(\theta) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(e^{\alpha W}y - X\beta)' e^{\tau M'} e^{\tau M}(e^{\alpha W}y - X\beta), \tag{2.2}$$

where $\theta = (\alpha, \tau, \beta', \sigma^2)'$. Note that (2.2) does not include the Jacobian terms, because $\ln|e^{\alpha W}| = \ln\left(e^{\alpha \mathrm{tr}(W)}\right) = 0$ and $\ln|e^{\tau M}| = \ln\left(e^{\tau \mathrm{tr}(M)}\right) = 0$, where $|\cdot|$ denotes the determinant operator, and $\mathrm{tr}(\cdot)$ is the trace operator. Let $\psi = (\alpha, \tau)'$. Then, for a given value of ψ, the first-order conditions of (2.2) with respect to β and σ yield

$$\widehat{\beta}(\psi) = (X' e^{\tau M'} e^{\tau M} X)^{-1} X' e^{\tau M'} e^{\tau M} e^{\alpha W} y, \tag{2.3}$$

$$\widehat{\sigma}^2(\psi) = \frac{1}{n}(e^{\alpha W}y - X\widehat{\beta})' e^{\tau M'} e^{\tau M}(e^{\alpha W}y - X\widehat{\beta}) = \frac{1}{n} y' e^{\alpha W'} e^{\tau M'} H(\tau) e^{\tau M} e^{\alpha W} y, \tag{2.4}$$

where $H(\tau) = I_n - e^{\tau M} X (X' e^{\tau M'} e^{\tau M} X)^{-1} X' e^{\tau M'}$. Then, ignoring constant terms, the concentrated likelihood function can be written as

$$\ln L^c(\theta) = -\ln(\widehat{\sigma}^2(\psi)) = -\ln\left(\frac{1}{n} y' e^{\alpha W'} e^{\tau M'} H(\tau) e^{\tau M} e^{\alpha W} y\right). \tag{2.5}$$

Thus, we can define the QMLE of ψ_0 as

$$\widehat{\psi} = \mathrm{argmin}_{\psi}\left(y' e^{\alpha W'} e^{\tau M'} H(\tau) e^{\tau M} e^{\alpha W} y\right). \tag{2.6}$$

Let $\gamma = (\alpha, \tau, \beta')'$ and $\gamma_0 = (\alpha_0, \tau_0, \beta_0')'$ be the true parameter vector. Then, QMLE $\widehat{\gamma}$ has the following asymptotic normal distribution (Debarsy et al., 2015),[5]

$$\sqrt{n}\left(\widehat{\gamma} - \gamma_0\right) \xrightarrow{d} N\left(0, \lim_{n\to\infty} C^{-1}\Omega C^{-1}\right). \tag{2.7}$$

Here, $\Omega = 2\sigma_0^2 C + \Omega_1$ with

$$C = \frac{1}{n}\begin{pmatrix} \sigma_0^2 \,\mathrm{tr}\left(\mathbb{W}^s \mathbb{W}^s\right) + 2\left(\mathbb{W} e^{\tau_0 M} X\beta_0\right)'\left(\mathbb{W} e^{\tau_0 M} X\beta_0\right) & * & * \\ \sigma_0^2 \,\mathrm{tr}\left(\mathbb{W}^s M^s\right) & \sigma_0^2 \,\mathrm{tr}\left(M^s M^s\right) & * \\ -2\left(e^{\tau_0 M} X\right)' \mathbb{W} e^{\tau_0 M} X\beta_0 & 0 & 2\left(e^{\tau_0 M} X\right)'\left(e^{\tau_0 M} X\right) \end{pmatrix}, \tag{2.8}$$

[5] Note that under heteroskedasticity of an unknown form, the QMLE is still consistent and has an asymptotically normal distribution when W and M commute, i.e., $WM = MW$ (Debarsy et al., 2015).

$$\Omega_1 = \frac{1}{n}\begin{pmatrix} \left(\mu_4 - 3\sigma_0^4\right)\mathrm{vec}_{\mathrm{D}}'\left(\mathbb{W}^s\right)\mathrm{vec}_{\mathrm{D}}\left(\mathbb{W}^s\right) + 4\mu_3\left(\mathbb{W}e^{\tau_0 M}X\beta_0\right)'\mathrm{vec}_D\left(\mathbb{W}^s\right) & * & * \\ 0 & 0 & * \\ -2\mu_3\left(e^{\tau_0 M}X\right)'\mathrm{vec}_{\mathrm{D}}\left(\mathbb{W}^s\right) & 0 & 0 \end{pmatrix}, \tag{2.9}$$

where $\mathbb{W} = e^{\tau_0 M}We^{-\tau_0 M}$, $\mu_3 = \mathrm{E}\epsilon_i^3$, $\mu_4 = \mathrm{E}\epsilon_i^4$, $\mathrm{vec}_{\mathrm{D}}(A)$ denotes a vector containing the diagonal elements of any square matrix A, and $B^s = B + B'$ for any $n \times n$ matrix B.

2.3 GMM Approach

When the error terms are homoskedastic, a set of moment functions that consists of linear and quadratic moment functions can be arranged such that the resulting GMME is as efficient as the MLE under the normal case, and asymptotically more efficient than the QMLE under the non-normal case. In the case of heteroskedasticity of an unknown form, an optimal GMME (OGMME) can be defined such that it is also more efficient than the QMLE (Debarsy et al., 2015).

To define the best set of moment functions when the error terms are simply i.i.d, we introduce the following notations. Let $\mathrm{Diag}(a)$ be the $n \times n$ matrix whose diagonal entries are the elements of the $n \times 1$ vector a, and $\mathrm{Diag}(A)$ be the $n \times n$ diagonal matrix whose diagonal entries are those of the $n \times n$ matrix A. Let $B^{(t)} = B - I_n\mathrm{tr}(B)/n$ for any $n \times n$ matrix B, and X_n^* be the submatrix of X with the intercept term removed. Define $P_1^* = \mathbb{W}, P_2^* = \mathrm{Diag}(\mathbb{W})$, $P_3^* = \mathrm{Diag}\left(e^{\tau_0 M}WX\beta_0\right)^{(t)}$, $P_4^* = M$, $P_{l+4}^* = \mathrm{Diag}\left(e^{\tau_0 M}X_l^*\right)^{(t)}$ for $l = 1, \ldots, k^*$ and $F^* = (F_1^*, F_2^*, F_3^*, F_4^*)$ with $F_1^* = e^{\tau_0 M}X^*, F_2^* = e^{\tau_0 M}WX\beta_0, F_3^* = l_n$, and $F_4^* = \mathrm{vec}_D\left(\mathbb{W}\right)$. Then, under the assumption that the error terms are i.i.d, the best set of moment functions suggested in Debarsy et al. (2015) takes the following form:

$$g^*(\gamma) = \frac{1}{n}\left(\epsilon'(\gamma)P_1^*\epsilon(\gamma), \ldots, \epsilon'(\gamma)P_{k^*+4}^*\epsilon(\gamma), \epsilon'(\gamma)F^*\right)', \tag{2.10}$$

where $\epsilon(\gamma) = e^{\tau M}\left(e^{\alpha W}y - X\beta\right)$. Then, the best GMME (BGMME) is defined as

$$\widehat{\gamma}_B = \mathrm{argmin}_{\gamma}\, g^{*'}(\gamma)V^{*-1}g^*(\gamma), \tag{2.11}$$

where $V^* = n\mathrm{E}\left(g^*\left(\gamma_0\right)g^{*'}\left(\gamma_0\right)\right)$. It can be shown that

$$V^* = \frac{1}{n}\begin{pmatrix} \frac{\sigma_0^4}{2}\omega'\omega + \frac{1}{4}\left(\mu_4 - 3\sigma_0^4\right)\omega_d'\omega_d & \frac{1}{2}\mu_3\omega_d'F^* \\ \frac{1}{2}\mu_3 F^{*\prime}\omega_d & \sigma_0^2 F^{*\prime}F^* \end{pmatrix}, \tag{2.12}$$

where $\omega_d = \left(\mathrm{vec}_D\left(P_1^{*s}\right), \ldots \mathrm{vec}_{\mathrm{D}}\left(P_{k^*+4}^{*s}\right)\right)$, $\omega = \left(\mathrm{vec}\left(P_1^{*s}\right), \ldots, \mathrm{vec}(P_{k^*+4}^{*s})\right)$ and $\mathrm{vec}(A)$ denotes the vectorization of matrix A. Under some regularity conditions, it follows that (Debarsy et al., 2015)

$$\sqrt{n}\left(\widehat{\gamma}_B - \gamma_0\right) \xrightarrow{d} N\left(0, \lim_{n\to\infty}(G^{*'}V^{*-1}G^*)^{-1}\right), \tag{2.13}$$

where

$$G^* = \mathbb{E}\left(\frac{\partial g^*(\gamma_0)}{\partial \gamma'}\right) = \frac{1}{n}\left(\begin{array}{ccc} \frac{\sigma_0^2}{2}\omega' \operatorname{vec}(\mathbb{W}^s) & \frac{\sigma_0^2}{2}\omega' \operatorname{vec}(M^s) & 0 \\ F^{*\prime}\mathbb{W}e^{\tau_0 M}X\beta_0 & 0 & -F^{*\prime}e^{\tau_0 M}X \end{array}\right). \tag{2.14}$$

Note that the BGMME defined (2.11) is not feasible, since V^*, $P_1^*, \ldots, P_{k^*+4}^*$ and F^* are functions of the unknown parameters. In practice, an initial consistent estimator of γ_0 can be used to replace the unknown parameters in these terms.[6] A feasible estimator formulated in this way can be shown to have the the same asymptotic distribution as that of $\widehat{\gamma}_B$ (Debarsy et al., 2015).

When the disturbance terms have heteroskedasticity of an uknown form, the matrices P_i's used in the quadratic moment functions need to have zero diagonal elements. In the heteroskedastic case, an OGMME can be defined by using the following vector of moment functions,

$$\widehat{g}_d(\gamma) = \frac{1}{n}\left(\epsilon'(\gamma)\left(\widehat{\mathbb{W}} - \operatorname{Diag}(\widehat{\mathbb{W}})\right)\epsilon(\gamma), \epsilon'(\gamma)M\epsilon(\gamma), \epsilon'(\gamma)\left(\widehat{\mathbb{W}}e^{\widehat{\tau}M}X\widehat{\beta}, e^{\widehat{\tau}M}X\right)\right)', \tag{2.15}$$

where $\widehat{\mathbb{W}} = e^{\widehat{\tau}M}We^{-\widehat{\tau}M}$.

2.4 Bayesian Approach

To complete the model specification, we need to specify the prior distributions for α, τ, β and σ^2. We assume the following prior distributions: $\alpha \sim N(\mu_\alpha, V_\alpha)$, $\tau \sim N(\mu_\tau, V_\tau)$, $\beta \sim N(\mu_\beta, V_\beta)$, and $\sigma^2 \sim IG(a_0, b_0)$, where IG denotes the inverse-gamma distribution. The likelihood is given by

$$y|\alpha, \tau, \beta, \sigma^2 \sim N(e^{-\alpha W}X\beta, \sigma^2 e^{-\alpha W}e^{-\tau M}e^{-\tau M'}e^{-\alpha W'}). \tag{2.16}$$

The standard Bayesian analysis for a linear regression model can be used to obtain the conditional posterior distributions of β and σ^2. On the other hand, combining the likelihood function with the prior densities of spatial parameters indicates that the conditional posterior distributions of α and τ are non-standard. In the following algorithm, we suggest a Gibbs sampler that shows how to generate random draws from the joint posterior distribution $p(\beta, \sigma^2, \alpha, \tau|y)$.[7]

Algorithm 1.
1. Sampling step for β:

$$\beta|y, \alpha, \tau, \sigma^2 \sim N(\widehat{\beta}, K_\beta), \tag{2.17}$$

where $K_\beta = (V_\beta^{-1} + X'e^{\tau M'}e^{\tau M}X/\sigma^2)^{-1}$ and $\widehat{\beta} = K_\beta(X'e^{\tau M'}e^{\tau M}e^{\alpha W}y/\sigma^2 + V_\beta^{-1}\mu_\beta)$.

[6] Among other consistent estimators, the following initial GMME can be used: $\widehat{\gamma} = \operatorname{argmin}_\gamma g'(\gamma)g(\gamma)$, where $g(\gamma) = (W\epsilon(\gamma), M\epsilon(\gamma), WX, X)'\epsilon(\gamma)$ and $\epsilon(\gamma) = e^{\tau M}\left(e^{\alpha W}y - X\beta\right)$.

[7] We use $p(\cdot)$ to denote the relevant density function, and we omit X in the conditional sets for the simplicity of exposition.

2. Sampling step for σ^2:

$$\sigma^2|y,\alpha,\tau,\beta \sim IG(\widehat{\sigma}^2, K_{\sigma^2}), \tag{2.18}$$

where $\widehat{\sigma}^2 = a_0 + \frac{n}{2}$ *and* $K_{\sigma^2} = b_0 + \frac{1}{2}(e^{\alpha W}y - X\beta)' e^{\tau M'} e^{\tau M}(e^{\alpha W}y - X\beta)$.

3. Sampling step for α:

$$p(\alpha|y,\beta,\tau,\sigma^2) \propto \exp\left(-\frac{1}{2}\left(\sigma^{-2}(e^{\alpha W}y - X\beta)' e^{\tau M'} e^{\tau M}(e^{\alpha W}y - X\beta) + V_{\alpha}^{-1}(\alpha^2 - 2\mu_{\alpha}\alpha)\right)\right), \tag{2.19}$$

which is a non-standard distribution. We can use a random-walk Metropolis-Hastings algorithm to sample from this distribution (LeSage and Pace, 2009). A candidate value α^{new} is generated according to

$$\alpha^{new} = \alpha^{old} + c_{\alpha} \times N(0,1), \tag{2.20}$$

where c_{α} is the tuning parameter.[8] *The candidate value α^{new} is accepted with probability*

$$\mathbb{P}(\alpha^{new}, \alpha^{old}) = min\left(1, \frac{p(\alpha^{new}|y,\beta,\sigma^2,\tau)}{p(\alpha^{old}|y,\beta,\sigma^2,\tau)}\right). \tag{2.21}$$

4. Sampling step for τ:

$$p(\tau|y,\beta,\alpha,\sigma^2) \propto \exp\left(-\frac{1}{2}\left(\sigma^{-2}(e^{\alpha W}y - X\beta)' e^{\tau M'} e^{\tau M}(e^{\alpha W}y - X\beta) + V_{\tau}^{-1}(\tau^2 - 2\mu_{\tau}\tau)\right)\right). \tag{2.22}$$

We use the random-walk Metropolis-Hastings algorithm described in Step 3 to generate random draws from $p(\tau|y,\beta,\alpha,\sigma^2)$.

2.5 Impact Measures

The dispersion of parameter estimators can be estimated in different ways (Arbia et al., 2020; Debarsy et al., 2015; Elhorst, 2014; LeSage and Pace, 2009; Taşpınar et al., 2018). In the case of QMLE and GMME defined in Sections 2.2 and 2.3, the closed-forms of variance-covariance matrices are available. Thus, we can use the plug-in method for these estimators. That is, the unknown parameters in these variance-covariance matrices can be replaced by the corresponding estimates obtained from consistent estimators. In the case of Bayesian approach, we can use the empirical standard deviations of the random draws generated through Algorithm 1 as the estimate for the standard errors of parameters.

According to model in (2.1), the derivative of y with respect to the kth explanatory variable x_k gives the marginal effect $e^{-\alpha_0 W}\beta_{0k}$, where β_{0k} is the kth element of the true coefficient vector

[8] The tuning parameter is determined during the estimation such that the acceptance rate falls between 40% and 60%.

β_0. To ease the interpretation and presentation of this marginal effect, LeSage and Pace (2009) define three scalar measures for the marginal effect: the average direct impact, the average indirect impact, and the average total impact. The average direct impact is the average of the main diagonal elements of $e^{-\alpha_0 W}\beta_{0k}$, the average indirect impact is the average of the off-diagonal elements of $e^{-\alpha_0 W}\beta_{0k}$, and the total impact is the average of the all elements of $e^{-\alpha_0 W}\beta_{0k}$. For statistical inference, one needs to determine the dispersions of these scalar impact measures. In the Bayesian approach, a sequence of random draws obtained through Algorithm 1 can be used to generate a sequence of random draws for each impact measure. Then, the mean and the standard deviation calculated from each sequence of impact measures can be used for inference.

In the QML and GMM cases, the classical delta method can be used to determine the dispersions of the impact measures. The estimator of the average direct effect is given by $\frac{1}{n}\mathrm{tr}(e^{-\widehat{\alpha}W_n}\widehat{\beta}_k)$. Then, by the mean value theorem, we obtain

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\left(\mathrm{tr}(e^{-\widehat{\alpha}W}\widehat{\beta}_k)-\mathrm{tr}(e^{-\alpha_0 W}\beta_{0k})\right)\\
&=\frac{1}{\sqrt{n}}\left(-\mathrm{tr}(e^{-\widehat{\alpha}W}W\widehat{\beta}_k)(\widehat{\alpha}-\alpha_0)+\mathrm{tr}(e^{-\widehat{\alpha}W})(\widehat{\beta}_k-\beta_{0k})\right)+o_p(1)\\
&=A_1\times\sqrt{n}(\widehat{\alpha}-\alpha_0,\widehat{\beta}_k-\beta_{0k})^{'}+o_p(1)\xrightarrow{d}N(0,\lim_{n\to\infty}A_1BA_1^{'}),
\end{aligned}\tag{2.23}
$$

where $A_1=\left(-\frac{1}{n}\mathrm{tr}(e^{-\alpha W}W\beta_k),\frac{1}{n}\mathrm{tr}(e^{-\alpha W})\right)$, B is the asymptotic covariance of $\sqrt{n}(\widehat{\alpha}-\alpha_0,\widehat{\beta}_k-\beta_{0k})$. So the asymptotic variance of direct effects can be estimated by $\frac{1}{n}\widehat{A}_1\widehat{B}\widehat{A}_1^{'}$, where $\widehat{A}_1=\left(-\frac{1}{n}\mathrm{tr}(e^{-\widehat{\alpha}W}W\widehat{\beta}_k),\frac{1}{n}\mathrm{tr}(e^{-\widehat{\alpha}W})\right)$, and $\widehat{B}$ is the estimated asymptotic covariance of $\sqrt{n}(\widehat{\alpha}-\alpha_0,\widehat{\beta}_k-\beta_{0k})$. Applying the mean value theorem to the estimator of total effect $\frac{1}{n}\widehat{\beta}_k l_n^{'}e^{-\widehat{\alpha}W}l_n$, where l_n is the $n\times 1$ vector of ones, we obtain

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\left(\widehat{\beta}_k l_n^{'}e^{-\widehat{\alpha}W}l_n-\beta_{0k}l_n^{'}e^{-\alpha_0 W}l_n\right)\\
&=A_2\times\sqrt{n}(\widehat{\alpha}-\alpha_0,\widehat{\beta}_k-\beta_{0k})^{'}+o_p(1)\xrightarrow{d}N(0,\lim_{n\to\infty}A_2BA_2^{'}),
\end{aligned}\tag{2.24}
$$

where $A_2=\left(-\frac{1}{n}\beta_k l_n^{'}e^{-\alpha_0 W}Wl_n,\frac{1}{n}l_n^{'}e^{-\alpha_0 W}l_n\right)$. Thus, $\mathrm{Var}(\frac{1}{n}\widehat{\beta}_k l_n^{'}e^{-\widehat{\alpha}W}l_n)$ can be estimated by $\frac{1}{n}\widehat{A}_2\widehat{B}\widehat{A}_2^{'}$, where $\widehat{A}_2=\left(-\frac{1}{n}\widehat{\beta}_k l_n^{'}e^{-\widehat{\alpha}W}Wl_n,\frac{1}{n}l_n^{'}e^{-\widehat{\alpha}W}l_n\right)$. Finally, applying the mean value theorem to the estimator of average indirect effects $\frac{1}{n}\left(\widehat{\beta}_k l_n^{'}e^{-\widehat{\alpha}W}l_n-\mathrm{tr}(e^{-\widehat{\alpha}W}\widehat{\beta}_k)\right)$, we obtain

$$
\begin{aligned}
&\frac{1}{\sqrt{n}}\left(\left(\widehat{\beta}_k l_n^{'}e^{-\widehat{\alpha}W}l_n-\mathrm{tr}(e^{-\widehat{\alpha}W}\widehat{\beta}_k)\right)-\left(\beta_{0k}l_n^{'}e^{-\widehat{\alpha}_0 W}l_n-\mathrm{tr}(e^{-\widehat{\alpha}_0 W}\beta_{0k})\right)\right)=\\
&(A_2-A_1)\times\sqrt{n}(\widehat{\alpha}-\alpha_0,\widehat{\beta}_k-\beta_{0k})^{'}+o_p(1)\xrightarrow{d}N(0,\lim_{n\to\infty}(A_2-A_1)B(A_2-A_1)^{'}).
\end{aligned}\tag{2.25}
$$

Then, an estimate of $\mathrm{Var}\left(\frac{1}{n}\left(\widehat{\beta}_k l_n^{'}e^{-\widehat{\alpha}W}l_n-\mathrm{tr}(e^{-\widehat{\alpha}W}\widehat{\beta}_k)\right)\right)$ is given by $\frac{1}{n}(\widehat{A}_2-\widehat{A}_1)\widehat{B}(\widehat{A}_2-\widehat{A}_1)'$.

3 The matrix-vector products method

Our analysis in Section 2 indicates that the estimation of MESS$(1,1)$ specifically requires the evaluation of $e^{\tau M}e^{\alpha W}y$ and $e^{\tau M}X$. In the case of the QMLE, the objective function in (2.6) is comprised of $e^{\tau M}e^{\alpha W}y$ and the term $H(\tau)$, which is a function of $e^{\tau M}X$. For the BGMME in (2.11), the vector of best moment functions $g^*(\gamma)$ contains the disturbances $\epsilon(\gamma)$, which is a function of $e^{\tau M}e^{\alpha W}y$ and $e^{\tau M}X$. Our Algorithm 1 indicates that the Bayesian estimator also requires the evaluation of $e^{\tau M}e^{\alpha W}y$ and $e^{\tau M}X$ in each pass through the Gibbs sampler. In this section, we show how the matrix-vector products approach can be used to compute $e^{\tau M}e^{\alpha W}y$ and $e^{\tau M}X$ based on the truncation of Taylor series expansion of matrix exponential terms.

We start with $e^{\tau M}e^{\alpha W}y$. By definition, we have $e^{\tau M}e^{\alpha W}y = \sum_{i=0}^{\infty}\frac{\tau^i M^i}{i!}\sum_{j=0}^{\infty}\frac{\alpha^j W^j}{j!}y$. Truncating the Taylor series at the $(q+1)$th order yields $e^{\tau M}e^{\alpha W}y \approx \sum_{i=0}^{q}\frac{\tau^i M^i}{i!}\sum_{j=0}^{q}\frac{\alpha^j W^j}{j!}y$. Note that we can express $\sum_{i=0}^{q}\frac{\tau^i M^i}{i!}\sum_{j=0}^{q}\frac{\alpha^j W^j}{j!}$ as

$$\sum_{i=0}^{q}\frac{\tau^i M^i}{i!}\sum_{j=0}^{q}\frac{\alpha^j W^j}{j!} = \sum_{i=1}^{q}\sum_{j=0}^{i-1}\left(\frac{\tau^i\alpha^j M^i W^j}{i!j!} + \frac{\tau^j\alpha^i M^j W^i}{i!j!}\right) + \sum_{i=0}^{q}\frac{\tau^i\alpha^i M^i W^i}{(i!)^2}. \tag{3.1}$$

Let $\mathrm{Diag}(a_1, a_2, \ldots, a_n)$ be the $n\times n$ diagonal matrix with diagonal elements $\{a_1, a_2, \ldots, a_n\}$. Then, using (3.1), we can express $e^{\tau M}e^{\alpha W}y$ in the following way,

$$\begin{aligned} e^{\tau M}e^{\alpha W}y &\approx \sum_{i=1}^{q}\sum_{j=0}^{i-1}\frac{\tau^i\alpha^j M^i W^j}{i!j!}y + \sum_{i=1}^{q}\sum_{j=0}^{i-1}\frac{\tau^j\alpha^i M^j W^i}{i!j!}y + \sum_{i=0}^{q}\frac{\tau^i\alpha^i M^i W^i}{(i!)^2}y \\ &= Y_1 D_1 \nu_1(\alpha,\tau) + Y_2 D_2 \nu_2(\alpha,\tau) + Y_3 D_3 \nu_3(\alpha,\tau), \end{aligned} \tag{3.2}$$

where Y_1 and Y_2 are $n\times\frac{q(q+1)}{2}$ matrices, Y_3 is an $n\times(q+1)$ matrix, D_1 and D_2 are $\frac{q(q+1)}{2}\times\frac{q(q+1)}{2}$ matrices, D_3 is an $(q+1)\times(q+1)$ matrix, $\nu_1(\alpha,\tau)$ and $\nu_2(\alpha,\tau)$ are $\frac{q(q+1)}{2}\times 1$ vectors, and $\nu_3(\alpha,\tau)$ is an $(q+1)\times 1$ vector. The terms in (3.2) are

$$Y_1 = \left[My, M^2y, M^2Wy, M^3y, M^3Wy, M^3W^2y, \ldots, M^qy, M^qWy, \ldots, M^qW^{q-1}y\right], \tag{3.3}$$

$$Y_2 = \left[Wy, W^2y, MW^2y, W^3y, MW^3y, M^2W^3y, \ldots, W^qy, MW^qy, \ldots, M^{q-1}W^qy\right], \tag{3.4}$$

$$Y_3 = \left[y, MWy, M^2W^2y, \ldots, M^qW^qy\right], \tag{3.5}$$

$$D_1 = D_2 = \mathrm{Diag}\left(\frac{1}{0!1!}, \frac{1}{0!2!}, \frac{1}{1!2!}, \ldots, \frac{1}{0!q!}, \ldots, \frac{1}{(q-1)!q!}\right), \tag{3.6}$$

$$D_3 = \mathrm{Diag}\left(\frac{1}{(0!)^2}, \frac{1}{(1!)^2}, \frac{1}{(2!)^2}, \ldots, \frac{1}{(q!)^2}\right), \tag{3.7}$$

$$\nu_1(\alpha,\tau) = \left[\tau, \tau^2, \tau^2\alpha, \tau^3, \tau^3\alpha, \tau^3\alpha^2, \ldots, \tau^q, \tau^q\alpha, \ldots, \tau^q\alpha^{q-1}\right]', \tag{3.8}$$
$$\nu_2(\alpha,\tau) = \left[\alpha, \alpha^2, \alpha^2\tau, \alpha^3, \alpha^3\tau, \alpha^3\tau^2, \ldots, \alpha^q, \alpha^q\tau, \ldots, \alpha^q\tau^{q-1}\right]', \tag{3.9}$$
$$\nu_3(\alpha,\tau) = \left[1, \tau\alpha, \tau^2\alpha^2, \tau^3\alpha^3, \ldots, \tau^q\alpha^q\right]'. \tag{3.10}$$

Next, we show how the matrix-vector products approach can be used to get an approximation of $e^{\tau M}X$. Let $X = [X_1, X_2, \ldots, X_k]$, where X_i is the ith column. Then, we can write $e^{\tau M}X$ as

$$e^{\tau M}X = \left[e^{\tau M}X_1, e^{\tau M}X_2, \ldots, e^{\tau M}X_k\right] \approx \mathbb{X}D_4\mu(\tau), \tag{3.11}$$

where $\mathbb{X}$ is an $n \times k(q+1)$ matrix, D_4 is an $k(q+1) \times k(q+1)$ matrix and $\mu(\tau)$ is an $k(q+1) \times k$ matrix. It can be shown that

$$\mathbb{X} = \left[X_1, MX_1, M^2X_1, \ldots, M^qX_1, X_2, MX_2, M^2X_2, \ldots, M^qX_2, \ldots, X_k, MX_k, M^2X_k, \ldots,\ M^qX_k\right], \tag{3.12}$$

$$D_4 = I_k \otimes \mathrm{Diag}\left(\frac{1}{0!}, \frac{1}{1!}, \frac{1}{2!}, \ldots, \frac{1}{q!}\right), \tag{3.13}$$

$$\mu(\tau) = I_k \otimes \left[1, \tau, \tau^2, \ldots, \tau^q\right]'. \tag{3.14}$$

Next, we discuss how the matrix-vector products approach can be applied to the estimators given in Section 2. Recall that the QMLE is defined by $\widehat{\psi} = \mathrm{argmin}_{\psi}\left(y'e^{\alpha W'}e^{\tau M'}H(\tau)e^{\tau M}e^{\alpha W}y\right)$, where the objective function requires the evaluation of $e^{\tau M}e^{\alpha W}y$ and $e^{\tau M}X$ in each iteration. Using the matrix-vector products approach, we can avoid the evaluation of these terms in each iteration of the optimization routine. We can define $\mathbb{X}$, Y_i, and D_j for $i \in \{1,2,3\}$ and $j \in \{1,2,3,4\}$, and then supply these terms as the inputs of the objective function in the optimization solver. In this way, we can avoid the computation of these terms in each iteration.

In the case of GMM approach, the vector of moment functions in (2.10) contains the disturbance terms $\epsilon(\gamma) = e^{\tau M}e^{\alpha W}y - e^{\tau M}X\beta$, which can also be expressed in the matrix-vector products approach by using (3.2) and (3.11). When implementing (2.11), similar to the case of QMLE, we can define $\mathbb{X}$, Y_i, and D_j for $i \in \{1,2,3\}$ and $j \in \{1,2,3,4\}$, and then supply these terms as inputs for the objective function to the optimization solver.

Finally, in the case of Bayesian approach, we need to work with expressions in Algorithm 1. Before implementing the Gibbs sampler described in Algorithm 1, we can compute the required terms for the matrix exponential terms and then pass these terms to the sampler. That is, we can define $\mathbb{X}$, Y_i, and D_j for $i \in \{1,2,3\}$ and $j \in \{1,2,3,4\}$, and then supply these terms to the Gibbs sampler. Thus, we can avoid computation of these terms in each pass of the Gibbs sampler.

Remark 1. *The matrix-vector products approach is general enough and can be easily adjusted for* *some other order MESS models. For example, the expressions for two special cases, MESS*$(1,0)$ *and MESS*$(0,1)$*, can be simply obtained by setting* M *and* W *to the zero matrix, respectively. Similarly, the Durbin versions of MESS*$(1,1)$*, MESS*$(1,0)$ *and MESS*$(0,1)$ *can also be estimated by defining independent variables appropriately. The matrix-vector products approach also applies to the MESS models with the panel data by arranging the terms involving the matrix exponential terms appropriately. For higher order MESS models, or MESS*(p,q)*, similar equations to* (3.1)-(3.14) *can be derived. For example, for the MESS*$(2,2)$*, we have*

$$\begin{aligned}
&e^{\tau_1 M_1+\tau_2 M_2}e^{\alpha_1 W_1+\alpha_2 W_2}y \approx \sum_{i=0}^{q}\frac{1}{i!}\left(\tau_1 M_1+\tau_2 M_2\right)^i \sum_{j=0}^{q}\frac{1}{j!}\left(\alpha_1 W_1+\alpha_2 W_2\right)^j y\\
&=\sum_{i=1}^{q}\sum_{j=0}^{i-1}\left(\frac{1}{i!j!}\sum_{k_1=0}^{i}\sum_{k_2=0}^{j}\binom{i}{k_1}\binom{j}{k_2}\tau_1^{i-k_1}\tau_2^{k_1}\alpha_1^{j-k_2}\alpha_2^{k_2}M_1^{i-k_1}M_2^{k_1}W_1^{j-k_2}W_2^{k_2}\right.\\
&\quad\left.+\frac{1}{i!j!}\sum_{k_1=0}^{j}\sum_{k_2=0}^{i}\binom{j}{k_1}\binom{i}{k_2}\tau_1^{j-k_1}\tau_2^{k_1}\alpha_1^{i-k_2}\alpha_2^{k_2}M_1^{j-k_1}M_2^{k_1}W_1^{i-k_2}W_2^{k_2}\right)\\
&\quad+\sum_{i=0}^{q}\sum_{k_1=0}^{i}\sum_{k_2=0}^{i}\binom{i}{k_1}\binom{i}{k_2}\tau_1^{i-k_1}\tau_2^{k_1}\alpha_1^{i-k_2}\alpha_2^{k_2}M_1^{i-k_1}M_2^{k_1}W_1^{i-k_2}W_2^{k_2}\\
&=\mathbb{Y}_1\mathbb{D}_1\omega_1(\alpha,\tau)+\mathbb{Y}_2\mathbb{D}_2\omega_2(\alpha,\tau)+\mathbb{Y}_3\mathbb{D}_3\omega_3(\alpha,\tau),
\end{aligned}\tag{3.15}$$

where $\mathbb{Y}_l$*,* $\mathbb{D}_l$ *and* ω_l *for* $l\in\{1,2,3\}$ *can be defined accordingly. In the case of* $e^{\tau_1 M_1+\tau_2 M_2}X$*, we have*

$$\begin{aligned}
&e^{\tau_1 M_1+\tau_2 M_2}X \approx \left[\sum_{i=0}^{q}\frac{1}{i!}(\tau_1 M_1+\tau_2 M_2)^i X_1,\ldots,\sum_{i=0}^{q}\frac{1}{i!}(\tau_1 M_1+\tau_2 M_2)^i X_k\right]\\
&=\left[\sum_{i=0}^{q}\frac{1}{i!}\sum_{j=0}^{i}\binom{i}{j}(\tau_1 M_1)^{i-j}(\tau_2 M_2)^j X_1,\ldots,\sum_{i=0}^{q}\frac{1}{i!}\sum_{j=0}^{i}\binom{i}{j}(\tau_1 M_1)^{i-j}(\tau_2 M_2)^j X_k\right]\\
&=\mathbb{X}\,\mathbb{D}_4\kappa(\tau),
\end{aligned}\tag{3.16}$$

where

$$\begin{aligned}
\mathbb{X}=\Big[&X_1, M_1X_1, M_2X_1, M_1^2X_1, M_1M_2X_1, M_2^2X_1,\ldots,M_1^qX_1, M_1^{q-1}M_2X_1,\ldots M_1M_2^{q-1}X_1, M_2^qX_1,\\
&\ldots,\\
&X_k, M_1X_k, M_2X_k, M_1^2X_k, M_1M_2X_k, M_2^2X_k,\ldots,M_1^qX_k, M_1^{q-1}M_2X_k,\ldots,M_1M_2^{q-1}X_k, M_2^qX_k\Big],
\end{aligned}$$

$$\mathbb{D}_4=I_k\otimes Diag\left(\frac{\binom{0}{0}}{0!},\frac{\binom{1}{0}}{1!},\frac{\binom{1}{1}}{1!},\ldots,\frac{\binom{q}{0}}{q!},\frac{\binom{q}{1}}{q!},\ldots,\frac{\binom{q}{q}}{q!}\right),$$

$$\kappa(\tau) = I_k \otimes \left[1, \tau_1, \tau_2, \tau_1^2, \tau_1\tau_2, \tau_2^2, \ldots, \tau_1^q, \ldots, \tau_1^{q-1}\tau_2, \ldots, \tau_1\tau_2^{q-1}, \tau_2^q\right]'.$$

Remark 2. *The matrix-vector products approach based on a large q value can provide a better accuracy, but it can increase the computation time. We may determine a satisfactory q value from the inverse error analysis for a scaling and squaring method with Taylor series approximation (the third method in Moler and Van Loan (1978, 2003)). Let $T_q(\alpha W) = \sum_{j=0}^{q} \frac{\alpha^j W^j}{j!}$. Since $e^{\alpha W} = \left(e^{2^{-m}\alpha W}\right)^{2^m}$, we may consider the approximation $\left(T_q(2^{-m}\alpha W)\right)^{2^m}$. By Corollary 1 in Moler and Van Loan (1978, 2003), if $\|2^{-m}\alpha W\|_\infty \leq 0.5$, then $\left(T_q(2^{-m}\alpha W)\right)^{2^m} = e^{\alpha W + E}$, where*

$$\frac{\|E\|_\infty}{\|\alpha W\|_\infty} \leq \left(\frac{1}{2}\right)^{q-3} \frac{1}{q+1} \approx \begin{cases} 1.5 \times 10^{-5} & \text{if} \quad q = 15 \\ 3.6 \times 10^{-7} & \text{if} \quad q = 20 \\ 2.4 \times 10^{-10} & \text{if} \quad q = 30 \end{cases}.$$

This inverse error analysis indicates that a value of $q = 15$ can be satisfactory.

4 Monte Carlo Simulations

In this section, we will investigate the implications of using the matrix-vector products method with the truncated Taylor series approximation versus the scaling and squaring algorithm with the Padé approximation (the `expm` function in MATLAB R2020b). To this end, we will explore the properties of the two competing methods in terms of computation time as well as their effects on the finite sample properties of the estimators described in Section 2.

We consider the following data generating process,

$$e^{\alpha W} y = \beta_1 X_1 + \beta_2 X_2 + u, \quad e^{\tau W} u = \epsilon, \tag{4.1}$$

where the elements of X_1 and X_2 are independently drawn from $U(0, \sqrt{12})$ and $N(0,1)$, respectively. For the spatial weights matrix $W = (w_{ij})$, we consider two cases, the rook contiguity and queen contiguity. 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. The weights matrices are then row normalized. We set $(\beta_1, \beta_2)' = (2, 1)'$, and let α and τ take values from $\{-2, -0.2, 0.2, 2\}$. The disturbance terms are generated according to $\epsilon_i \sim$ i.i.d. $N(0,1)$. We consider two sample sizes, $n = 169$ and $n = 361$. For the QMLE and GMME, we set the number of repetitions to 1000. In the case of Bayesian estimator, we set the number of repetitions to 100, the number of draws to 1500 and burn-ins to 500. We use the matrix-vector products approach with $q = 15$ in all cases.

We use the matrix-vector products method (denoted as mvp) and the scaling and squaring algorithm with the Padé approximation (denoted as expm) to obtain the parameter estimates, and the corresponding bias, root mean squared error (RMSE) and the coverage rate. We also compute the impact measures, which include the direct effect, the total effect and the indirect effect, and their respective bias, RMSE and empirical coverage. We report the total computation time in seconds over 1000 resamples.[9] In the case of the QMLE, the computation time for $\widehat{\psi} = (\widehat{\alpha}, \widehat{\tau})'$ includes the time to compute the estimates using the concentrated log-likelihood function in (2.5). The computation time for $\widehat{\beta}$ includes the time for $\widehat{\psi}$ and for $\widehat{\beta}$, which is computed using (2.3). The time for impact measures are consequently the sum of the computation time for $\widehat{\psi}$, $\widehat{\beta}$ and respective measures. In the case of BGMME, an initial GMM estimation is carried out to construct V^*, F^* and P^*'s. We use the following set of moments for the initial stage, $(W\epsilon(\gamma), M\epsilon(\gamma), WX, X)'\epsilon(\gamma)$. Thus, the computation time for the BGMME includes both stages for the estimation. The time for the impact measures are then computed by adding on corresponding computation time for impact measures. In the case of Bayesian estimation, the computation time includes the time for collecting 2000 draws including the burn-ins. Finally, the computation time for the impact measures are computed similarly to those in the cases of QMLE and GMME.

We focus on the simulation results provided in Tables 1–5.[10] Tables 1–3 report the simulation results for the QMLE case. The mvp method reduces the computation time by approximately 98% to 99% compared to the expm method for different sample sizes, while providing the same estimates for the parameters and the impact measures. In all cases, we obtain the same values for bias, RMSE and coverage rates under both methods. When $n = 169$, the computation time for the mvp approach is about 2% of the computation time for the expm method. For example, when $\alpha = -2$ and $\tau = -2$, the bias, RMSE and coverage of $\widehat{\alpha}$ using both methods are 0.003, 0.037 and 0.923, respectively. However, the computation time is 606.5 seconds for the expm method, and 11.8 seconds for the mvp method. This means that on average, each computation takes 0.6065 seconds using the expm method, and 0.0118 seconds using the mvp method. For $n = 361$, the bias, RMSE and coverage of $\widehat{\alpha}$ are again the same for both methods, but the computation time is 4168.4 seconds for the expm method, and 30.7 seconds for the mvp method, leading to an average running time of 4.1684 and 0.0307 seconds, respectively. These results show the matrix-vector products method reduces the computational burden significantly, while maintaining the finite sample properties of the estimators. Table 2 presents the results for $\widehat{\beta} = (\widehat{\beta}_1, \widehat{\beta}_2)'$. The findings are similar to those from Table 1. While providing a similar performance in term of the finite sample properties of the QMLE, the mvp method reduces the computation time by about 98% to 99% compared to the scaling and squaring algorithm. Table 3 presents the results for the average direct effect estimates for X_1 and X_2. There is no difference between the two competing methods in terms of bias, RMSE and coverage, but we again observe that the mvp method is computationally more efficient than the expm method.

Table 4 presents the results for $\widehat{\alpha}$ and $\widehat{\tau}$ in the GMM estimation case. The findings are very

[9] We use a MacBook Pro 2016 with 2.4GHz intel core i7 processor and 8 GB 1867 MHz LPDDR3 memory to run our simulations.

[10] Some additional results are given in Tables A.1–A.10 of Appendix A. The simulation results in these additional tables also attest that our suggested mvp method is computationally more efficient than the expm method.

similar to the findings in the QML estimation case. The computational time advantage of the mvp method is more prominent in the GMM case. The mvp method reduces the computation time by about 98% to 99% for $n = 169$ and 99.7% to 99.8% for $n = 361$ compared to the expm method. For example, when $n = 169$, the computation time for $\widehat{\gamma}$ is 1407.9 seconds using the expm method, and 60.6 seconds using the mvp method. It means that on average, one resample takes 1.4079 seconds to run for the expm method, and 0.0606 seconds to run for the mvp method. When $n = 361$, they are respectively 8663.3 seconds and 219.6 seconds, leading to an average time of 8.6633 seconds and 0.2196 seconds to run each resample.

Table 5 presents the results for $\widehat{\alpha}$ and $\widehat{\tau}$ in the Bayesian estimation case. The mvp method again provides the same computational performance over the expm method, while maintaining the same finite sample properties. It takes only about 1% when $n = 169$ and less than 1% when $n = 361$ of the total computation time for the expm method. For example, for $\alpha = -2$, $\tau = -2$ and $n = 169$, it takes 2779.4 seconds for the expm method, and 20.4 seconds for the mvp method to compute $\widehat{\alpha}$. Thus, on average it takes respectively 27.794 seconds and 0.204 seconds to collect draws for one resample. When $n = 361$, the computation time for $\widehat{\alpha}$ using the expm method increases by approximately sevenfold to 19741.3 seconds. But, the computation time for the mvp method only increases by about 20% to 25.9 seconds. Thus, it takes an average of 197.413 seconds to collect draws for each resample using the expm method, and 0.259 seconds using the mvp method.

Table 1: The QMLE results for $\widehat{\alpha}$ and $\widehat{\tau}$

		n=169								n=361							
		expm				mvp				expm				mvp			
		α				α				α				α			
		-2	-0.2	0.2	2	-2	-0.2	0.2	2	-2	-0.2	0.2	2	-2	-0.2	0.2	2
Results for $\widehat{\alpha}$																	
$\tau=-2$	Bias	.003	.003	.000	.002	.003	.003	.000	.002	$-.001$	.000	.000	.000	$-.001$	.000	.000	.000
	RMSE	.037	.034	.032	.035	.037	.034	.032	.035	.025	.023	.023	.024	.025	.023	.023	.024
	Coverage	.923	.934	.939	.945	.923	.934	.939	.945	.950	.954	.957	.938	.950	.954	.957	.938
	Time	606.5	607.9	591.1	619.5	11.8	11.9	11.6	11.6	4168.4	4264.6	4146.2	4228.9	30.7	38.2	37.2	30.8
$\tau=-0.2$	Bias	.000	.004	.001	.004	.000	.004	.001	.004	.000	.001	.000	.000	.000	.001	.000	.000
	RMSE	.058	.053	.055	.060	.058	.053	.055	.060	.041	.039	.037	.038	.041	.039	.037	.038
	Coverage	.927	.945	.943	.936	.927	.945	.943	.936	.937	.945	.944	.956	.937	.945	.944	.956
	Time	637.1	553.0	572.2	661.2	14.0	12.6	12.9	14.3	4149.6	3547.5	3670.0	4225.0	37.2	38.3	39.8	37.8
$\tau=0.2$	Bias	.002	.002	.002	.002	.002	.002	.002	.002	.001	$-.001$	.001	$-.001$	.001	$-.001$	.001	$-.001$
	RMSE	.055	.052	.053	.063	.055	.052	.053	.063	.040	.040	.041	.038	.040	.040	.041	.038
	Coverage	.931	.934	.947	.931	.931	.934	.947	.931	.946	.956	.925	.951	.946	.956	.925	.951
	Time	591.1	534.6	549.6	624.3	12.9	12.2	12.4	13.4	3838.9	3454.7	3498.7	4008.7	34.1	37.2	37.7	35.6
$\tau=2$	Bias	$-.001$	.000	.001	$-.001$	$-.001$	.000	.001	$-.001$	.000	$-.001$	.000	$-.001$	.000	$-.001$	.000	$-.001$
	RMSE	.024	.022	.027	.029	.024	.022	.027	.029	.017	.018	.021	.016	.017	.018	.021	.016
	Coverage	.905	.935	.936	.924	.905	.935	.936	.924	.934	.942	.942	.942	.934	.942	.942	.942
	Time	632.0	651.5	612.5	643.7	11.8	12.6	11.7	12.0	4355.0	4512.8	4270.1	4512.5	31.3	40.3	38.3	32.3
Results for $\widehat{\tau}$																	
$\tau=-2$	Bias	$-.012$	$-.003$	$-.008$	$-.011$	$-.012$	$-.003$	$-.008$	$-.011$	$-.006$	$-.007$	$-.007$	$-.008$	$-.006$	$-.007$	$-.007$	$-.008$
	RMSE	.146	.145	.147	.147	.146	.145	.147	.147	.104	.102	.102	.101	.104	.102	.102	.101
	Coverage	.948	.940	.941	.945	.948	.940	.941	.945	.940	.953	.945	.944	.940	.953	.945	.944
	Time	606.5	607.9	591.1	619.5	11.8	11.9	11.6	11.6	4168.4	4264.6	4146.2	4228.9	30.7	38.2	37.2	30.8
$\tau=-0.2$	Bias	$-.002$	.014	.004	.015	$-.002$	.014	.004	.015	.003	.001	.005	.001	.003	.001	.005	.001
	RMSE	.152	.149	.144	.151	.152	.149	.144	.151	.102	.103	.102	.101	.102	.103	.102	.101
	Coverage	.944	.938	.947	.927	.944	.938	.947	.927	.938	.937	.948	.948	.938	.937	.948	.948
	Time	637.1	553.0	572.2	661.2	14.0	12.6	12.9	14.3	4149.6	3547.5	3670.0	4225.0	37.2	38.3	39.8	37.8
$\tau=0.2$	Bias	.010	.015	.014	.011	.010	.015	.014	.011	.011	.007	.015	.004	.011	.007	.015	.004
	RMSE	.146	.149	.157	.154	.146	.149	.157	.154	.101	.104	.101	.105	.101	.104	.101	.105
	Coverage	.937	.948	.929	.938	.937	.948	.929	.938	.951	.939	.940	.948	.951	.939	.940	.948
	Time	591.1	534.6	549.6	624.3	12.9	12.2	12.4	13.4	3838.9	3454.7	3498.7	4008.7	34.1	37.2	37.7	35.6
$\tau=2$	Bias	.066	.059	.057	.055	.066	.059	.057	.055	.028	.032	.024	.026	.028	.032	.024	.026
	RMSE	.167	.165	.170	.161	.167	.165	.170	.161	.107	.106	.106	.105	.107	.106	.106	.105
	Coverage	.904	.908	.900	.916	.904	.908	.900	.916	.934	.936	.936	.941	.934	.936	.936	.941
	Time	632.0	651.5	612.5	643.7	11.8	12.6	11.7	12.0	4355.0	4512.8	4270.1	4512.5	31.3	40.3	38.3	32.3

Table 2: The QMLE results for $\widehat{\beta}_1$ and $\widehat{\beta}_2$

		n=169								n=361							
		expm				mvp				expm				mvp			
		α				α				α				α			
		-2	-0.2	0.2	2	-2	-0.2	0.2	2	-2	-0.2	0.2	2	-2	-0.2	0.2	2
Results for $\widehat{\beta}_1$																	
$\tau=-2$	Bias	.004	.003	$-.004$	$-.002$	.004	.003	$-.004$	$-.002$	.001	.003	.001	.000	.001	.003	.001	.000
	RMSE	.059	.060	.056	.057	.059	.060	.056	.057	.039	.038	.036	.035	.039	.038	.036	.035
	Coverage	.942	.937	.949	.943	.942	.937	.949	.943	.952	.944	.946	.952	.952	.944	.946	.952
	Time	610.3	611.2	594.5	623.2	12.0	12.1	11.7	11.8	4193.3	4287.2	4168.7	4253.5	31.1	38.7	37.7	31.2
$\tau=-0.2$	Bias	.002	.002	$-.002$	$-.001$	.002	.002	$-.002$	$-.001$	$-.002$	.000	$-.004$	.001	$-.002$	.000	$-.004$	.001
	RMSE	.079	.076	.074	.083	.079	.076	.074	.083	.051	.052	.054	.053	.051	.052	.054	.053
	Coverage	.937	.939	.955	.936	.937	.939	.955	.936	.947	.935	.931	.941	.947	.935	.931	.941
	Time	640.5	556.1	575.4	664.7	14.1	12.7	13.0	14.5	4171.9	3567.4	3689.9	4247.2	37.6	38.7	40.2	38.2
$\tau=0.2$	Bias	$-.001$	$-.001$	$-.002$	$-.003$	$-.001$	$-.001$	$-.002$	$-.003$	.002	.000	.000	.000	.002	.000	.000	.000
	RMSE	.084	.075	.079	.080	.084	.075	.079	.080	.053	.053	.054	.055	.053	.053	.054	.055
	Coverage	.938	.934	.942	.946	.938	.934	.942	.946	.951	.948	.936	.946	.951	.948	.936	.946
	Time	594.5	537.7	552.7	627.8	13.1	12.4	12.6	13.6	3861.0	3474.6	3518.6	4030.8	34.5	37.7	38.1	36.0
$\tau=2$	Bias	.002	.000	$-.002$	.000	.002	.000	$-.002$	.000	.001	$-.002$	$-.002$	.000	.001	$-.002$	$-.002$	.000
	RMSE	.036	.039	.036	.045	.036	.039	.036	.045	.028	.025	.029	.026	.028	.025	.029	.026
	Coverage	.921	.926	.940	.946	.921	.926	.940	.946	.943	.943	.944	.940	.943	.943	.944	.940
	Time	635.6	654.9	615.9	647.4	12.0	12.8	11.9	12.1	4379.6	4535.5	4292.8	4537.3	31.7	40.8	38.7	32.7
Results for $\widehat{\beta}_2$																	
$\tau=-2$	Bias	.001	.000	$-.001$	.002	.001	.000	$-.001$	.002	.001	$-.001$	$-.002$	.000	.001	$-.001$	$-.002$	.000
	RMSE	.056	.051	.048	.060	.056	.051	.048	.060	.039	.037	.037	.041	.039	.037	.037	.041
	Coverage	.946	.955	.934	.946	.946	.955	.934	.946	.948	.957	.945	.937	.948	.957	.945	.937
	Time	610.3	611.2	594.5	623.2	12.0	12.1	11.7	11.8	4193.3	4287.2	4168.7	4253.5	31.1	38.7	37.7	31.2
$\tau=-0.2$	Bias	.000	.000	$-.002$	$-.001$	.000	.000	$-.002$	$-.001$	.002	$-.002$	$-.004$	.000	.002	$-.002$	$-.004$	.000
	RMSE	.080	.075	.076	.075	.080	.075	.076	.075	.054	.052	.051	.051	.054	.052	.051	.051
	Coverage	.928	.938	.955	.953	.928	.938	.955	.953	.942	.949	.941	.945	.942	.949	.941	.945
	Time	640.5	556.1	575.4	664.7	14.1	12.7	13.0	14.5	4171.9	3567.4	3689.9	4247.2	37.6	38.7	40.2	38.2
$\tau=0.2$	Bias	.001	.002	$-.005$	$-.003$	.001	.002	$-.005$	$-.003$	.001	.002	.001	.002	.001	.002	.001	.002
	RMSE	.077	.086	.079	.090	.077	.086	.079	.090	.053	.053	.052	.052	.053	.053	.052	.052
	Coverage	.937	.929	.937	.944	.937	.929	.937	.944	.955	.941	.943	.946	.955	.941	.943	.946
	Time	594.5	537.7	552.7	627.8	13.1	12.4	12.6	13.6	3861.0	3474.6	3518.6	4030.8	34.5	37.7	38.1	36.0
$\tau=2$	Bias	.000	.001	$-.001$	$-.001$	.000	.001	$-.001$	$-.001$	$-.001$	.002	.000	.000	$-.001$	.002	.000	.000
	RMSE	.034	.034	.042	.036	.034	.034	.042	.036	.022	.025	.024	.027	.022	.025	.024	.027
	Coverage	.933	.941	.941	.926	.933	.941	.941	.926	.946	.941	.941	.931	.946	.941	.941	.931
	Time	635.6	654.9	615.9	647.4	12.0	12.8	11.9	12.1	4379.6	4535.5	4292.8	4537.3	31.7	40.8	38.7	32.7

Table 3: The QMLE results of direct effects for X_1 and X_2

		n=169								n=361							
		expm				mvp				expm				mvp			
		α				α				α				α			
		-2	-0.2	0.2	2	-2	-0.2	0.2	2	-2	-0.2	0.2	2	-2	-0.2	0.2	2
Results for X_1																	
$\tau=-2$	Bias	.003	.003	$-.004$	.000	.003	.003	$-.004$	.000	.003	.003	.001	$-.001$	.003	.003	.001	$-.001$
	RMSE	.106	.060	.056	.109	.106	.060	.056	.109	.076	.039	.037	.070	.076	.039	.037	.070
	Coverage	.947	.937	.952	.945	.947	.937	.952	.945	.946	.941	.944	.946	.946	.941	.944	.946
	Time	612.0	612.6	595.9	625.0	13.8	13.5	13.2	13.6	4205.5	4297.0	4178.6	4265.6	43.4	48.5	47.5	43.3
$\tau=-0.2$	Bias	.005	.003	$-.001$	.007	.005	.003	$-.001$	.007	$-.001$	.000	$-.004$	.003	$-.001$	.000	$-.004$	.003
	RMSE	.159	.077	.075	.153	.159	.077	.075	.153	.105	.052	.054	.101	.105	.052	.054	.101
	Coverage	.948	.941	.961	.937	.948	.941	.961	.937	.954	.939	.932	.940	.954	.939	.932	.940
	Time	642.2	557.5	576.8	666.5	15.9	14.2	14.5	16.3	4184.3	3577.2	3699.8	4259.6	50.0	48.6	50.1	50.6
$\tau=0.2$	Bias	$-.003$	.000	$-.001$	.001	$-.003$	.000	$-.001$	.001	.003	.000	.000	.000	.003	.000	.000	.000
	RMSE	.163	.074	.080	.163	.163	.074	.080	.163	.106	.053	.055	.104	.106	.053	.055	.104
	Coverage	.940	.937	.942	.944	.940	.937	.942	.944	.949	.944	.941	.940	.949	.944	.941	.940
	Time	596.2	539.1	554.1	629.5	14.8	13.8	14.1	15.4	3873.4	3484.5	3528.4	4043.1	46.8	47.5	48.0	48.3
$\tau=2$	Bias	.005	.000	$-.001$	.000	.005	.000	$-.001$	.000	.003	$-.002$	$-.002$	.000	.003	$-.002$	$-.002$	.000
	RMSE	.079	.039	.037	.084	.079	.039	.037	.084	.055	.025	.030	.058	.055	.025	.030	.058
	Coverage	.922	.922	.946	.947	.922	.922	.946	.947	.930	.945	.939	.941	.930	.945	.939	.941
	Time	637.3	656.3	617.3	649.2	13.7	14.3	13.3	13.9	4391.6	4545.3	4302.7	4549.4	43.6	50.6	48.6	44.8
Results for X_2																	
$\tau=-2$	Bias	.000	.000	$-.001$	.006	.000	.000	$-.001$	.006	.002	$-.001$	$-.002$	.000	.002	$-.001$	$-.002$	.000
	RMSE	.096	.051	.048	.107	.096	.051	.048	.107	.068	.037	.038	.072	.068	.037	.038	.072
	Coverage	.942	.955	.936	.953	.942	.955	.936	.953	.943	.956	.945	.935	.943	.956	.945	.935
	Time	612.0	612.7	595.9	625.0	13.8	13.5	13.1	13.6	4205.5	4297.0	4178.6	4265.6	43.4	48.5	47.5	43.3
$\tau=-0.2$	Bias	.001	.000	$-.002$	.003	.001	.000	$-.002$	.003	.004	$-.002$	$-.004$	.000	.004	$-.002$	$-.004$	.000
	RMSE	.144	.076	.076	.126	.144	.076	.076	.126	.095	.053	.051	.088	.095	.053	.051	.088
	Coverage	.934	.938	.952	.950	.934	.938	.952	.950	.948	.948	.939	.933	.948	.948	.939	.933
	Time	642.2	557.5	576.8	666.5	15.9	14.2	14.5	16.2	4184.3	3577.2	3699.8	4259.6	50.0	48.6	50.1	50.6
$\tau=0.2$	Bias	.000	.002	$-.005$	$-.002$	.000	.002	$-.005$	$-.002$	.001	.002	.001	.003	.001	.002	.001	.003
	RMSE	.132	.086	.080	.156	.132	.086	.080	.156	.092	.053	.052	.089	.092	.053	.052	.089
	Coverage	.938	.930	.939	.940	.938	.930	.939	.940	.956	.942	.943	.952	.956	.942	.943	.952
	Time	596.2	539.1	554.1	629.5	14.8	13.8	14.1	15.4	3873.4	3484.5	3528.4	4043.1	46.9	47.5	48.0	48.3
$\tau=2$	Bias	.001	.001	$-.001$	$-.003$	.001	.001	$-.001$	$-.003$	$-.001$	.002	.000	$-.001$	$-.001$	.002	.000	$-.001$
	RMSE	.068	.034	.042	.061	.068	.034	.042	.061	.039	.025	.024	.045	.039	.025	.024	.045
	Coverage	.928	.937	.940	.939	.928	.937	.940	.939	.946	.941	.943	.936	.946	.941	.943	.936
	Time	637.4	656.3	617.3	649.2	13.7	14.3	13.3	13.9	4391.6	4545.3	4302.7	4549.4	43.7	50.6	48.6	44.7

Table 4: The GMME results for $\widehat{\alpha}$ and $\widehat{\tau}$

		n=169								n=361							
		expm				mvp				expm				mvp			
		α				α				α				α			
		−2	−0.2	0.2	2	−2	−0.2	0.2	2	−2	−0.2	0.2	2	−2	−0.2	0.2	2
Results for $\widehat{\alpha}$																	
τ=−2	Bias	.002	.002	−.001	.001	.002	.002	−.001	.001	−.001	.000	.000	−.001	−.001	.000	.000	−.001
	RMSE	.039	.035	.034	.037	.039	.035	.034	.037	.026	.023	.024	.024	.026	.023	.024	.024
	Coverage	.919	.930	.916	.920	.919	.930	.916	.920	.939	.947	.946	.930	.939	.947	.946	.930
	Time	1407.9	1298.5	1257.3	1270.5	60.6	57.0	56.2	54.9	8663.3	8788.2	8576.1	8797.2	219.6	220.6	219.6	219.7
τ=−0.2	Bias	−.001	.002	.000	.002	−.001	.002	.000	.002	−.002	.001	−.001	−.001	−.002	.001	−.001	−.001
	RMSE	.058	.056	.057	.062	.058	.056	.057	.062	.042	.040	.038	.039	.042	.040	.038	.039
	Coverage	.924	.931	.925	.914	.924	.931	.925	.914	.934	.945	.936	.942	.934	.945	.936	.942
	Time	1512.6	1413.2	1378.5	1504.4	60.8	59.7	58.8	60.5	9794.5	8872.9	8787.4	9936.7	207.6	203.1	202.2	208.4
τ=0.2	Bias	.001	.001	.001	.000	.001	.001	.001	.000	.000	−.002	.000	−.002	.000	−.002	.000	−.002
	RMSE	.057	.055	.054	.065	.057	.055	.054	.065	.040	.041	.042	.038	.040	.041	.042	.038
	Coverage	.921	.920	.935	.911	.921	.920	.935	.911	.945	.943	.925	.942	.945	.943	.925	.942
	Time	1409.1	1312.1	1300.7	1409.1	58.5	57.5	57.2	58.6	9341.3	8504.4	8409.4	9463.1	205.9	201.4	200.1	205.8
τ=2	Bias	−.002	−.001	.000	−.002	−.002	−.001	.000	−.002	.000	−.001	−.001	−.001	.000	−.001	−.001	−.001
	RMSE	.025	.023	.029	.030	.025	.023	.029	.030	.017	.018	.021	.017	.017	.018	.021	.017
	Coverage	.902	.932	.923	.913	.902	.932	.923	.913	.933	.937	.940	.935	.933	.937	.940	.935
	Time	1420.2	1335.1	1340.9	1343.2	58.4	58.2	58.4	56.4	9321.8	9157.7	8956.1	9167.3	223.0	224.4	222.9	221.0
Results for $\widehat{\tau}$																	
τ=−2	Bias	−.024	−.014	−.022	−.025	−.024	−.014	−.022	−.025	−.012	−.014	−.014	−.016	−.012	−.014	−.014	−.016
	RMSE	.151	.150	.152	.151	.151	.150	.152	.151	.105	.103	.105	.102	.105	.103	.105	.102
	Coverage	.941	.939	.930	.936	.941	.939	.930	.936	.937	.954	.934	.944	.937	.954	.934	.944
	Time	1407.9	1298.5	1257.3	1270.5	60.6	57.0	56.2	54.9	8663.3	8788.2	8576.1	8797.2	219.6	220.6	219.6	219.7
τ=−0.2	Bias	−.015	.001	−.011	.001	−.015	.001	−.011	.001	−.003	−.007	−.002	−.008	−.003	−.007	−.002	−.008
	RMSE	.154	.152	.148	.154	.154	.152	.148	.154	.104	.105	.102	.103	.104	.105	.102	.103
	Coverage	.936	.937	.936	.926	.936	.937	.936	.926	.935	.932	.938	.947	.935	.932	.938	.947
	Time	1512.6	1413.2	1378.5	1504.4	60.8	59.7	58.8	60.5	9794.5	8872.9	8787.4	9936.7	207.6	203.1	202.2	208.4
τ=0.2	Bias	−.006	−.003	−.005	−.006	−.006	−.003	−.005	−.006	.003	.000	.007	−.004	.003	.000	.007	−.004
	RMSE	.151	.151	.161	.158	.151	.151	.161	.158	.104	.105	.103	.106	.104	.105	.103	.106
	Coverage	.929	.942	.912	.921	.929	.942	.912	.921	.946	.926	.935	.938	.946	.926	.935	.938
	Time	1409.1	1312.1	1300.7	1409.1	58.5	57.5	57.2	58.6	9341.3	8504.4	8409.4	9463.1	205.9	201.4	200.1	205.8
τ=2	Bias	.039	.034	.030	.029	.039	.034	.030	.029	.016	.020	.011	.015	.016	.020	.011	.015
	RMSE	.164	.162	.167	.160	.164	.162	.167	.160	.107	.104	.105	.104	.107	.104	.105	.104
	Coverage	.907	.909	.898	.924	.907	.909	.898	.924	.935	.949	.940	.948	.935	.949	.940	.948
	Time	1420.2	1335.1	1340.9	1343.2	58.4	58.2	58.4	56.4	9321.8	9157.7	8956.1	9167.3	223.0	224.4	222.9	221.0

Table 5: The Bayesian results for $\widehat{\alpha}$ and $\widehat{\tau}$

		n=169								n=361							
		expm				mvp				expm				mvp			
		α				α				α				α			
		-2	-0.2	0.2	2	-2	-0.2	0.2	2	-2	-0.2	0.2	2	-2	-0.2	0.2	2
Results for $\widehat{\alpha}$																	
τ=-2	Bias	.006	$-.004$	.005	$-.002$	.006	$-.004$	.005	$-.002$	$-.002$	$-.003$	.001	.006	$-.002$	$-.003$	.001	.006
	RMSE	.036	.034	.038	.031	.036	.034	.038	.031	.028	.024	.022	.026	.028	.024	.022	.026
	Coverage	.960	.950	.940	.940	.960	.950	.940	.940	.950	.990	.980	.920	.950	.990	.980	.920
	Time	2779.4	2519.4	2522.4	2768.3	20.4	20.5	20.4	19.8	19741.3	17773.8	17742.6	19743.5	25.9	25.8	26.1	25.4
τ=-0.2	Bias	.001	.002	.002	$-.004$	.001	.002	.002	$-.004$	.012	.003	$-.001$	$-.008$	.012	.003	$-.001$	$-.008$
	RMSE	.059	.056	.050	.053	.059	.056	.050	.053	.039	.039	.035	.037	.039	.039	.035	.037
	Coverage	.950	.960	.950	.930	.950	.960	.950	.930	.900	.970	.960	.960	.900	.970	.960	.960
	Time	2569.1	2319.2	2319.2	2584.2	20.2	20.3	20.3	19.7	17645.8	15784.0	15748.3	17733.8	25.9	25.7	26.1	25.4
τ=0.2	Bias	.014	$-.005$	.003	$-.008$	.014	$-.005$	.003	$-.008$	.008	$-.005$	.001	$-.002$	.008	$-.005$	.001	$-.002$
	RMSE	.062	.051	.056	.065	.062	.051	.056	.065	.040	.041	.036	.042	.040	.041	.036	.042
	Coverage	.920	.970	.960	.940	.920	.970	.960	.940	.910	.950	.950	.970	.910	.950	.950	.970
	Time	2567.2	2333.8	2318.0	2570.9	20.3	20.0	20.1	20.1	17658.5	15786.3	15735.4	17797.9	26.1	25.4	25.8	25.2
τ=2	Bias	.000	.004	.002	$-.002$	.000	.004	.002	$-.002$	.000	.003	.000	$-.002$	.000	.003	.000	$-.002$
	RMSE	.025	.029	.035	.032	.025	.029	.035	.032	.023	.020	.020	.017	.023	.020	.020	.017
	Coverage	.960	.970	.960	.940	.960	.970	.960	.940	.930	.960	.940	.970	.930	.960	.940	.970
	Time	2771.2	2511.7	2526.1	2778.6	20.2	20.1	19.8	19.6	19692.2	17740.2	17745.2	19558.5	26.0	25.6	25.5	25.3
Results for $\widehat{\tau}$																	
τ=-2	Bias	.055	.050	.038	.073	.055	.050	.038	.073	.010	.029	.030	.022	.010	.029	.030	.022
	RMSE	.168	.157	.142	.155	.168	.157	.142	.155	.106	.111	.107	.105	.106	.111	.107	.105
	Coverage	.930	.920	.960	.920	.930	.920	.960	.920	.930	.890	.920	.920	.930	.890	.920	.920
	Time	2768.4	2512.7	2515.5	2761.3	37.5	37.7	37.5	36.6	19666.9	17777.3	17735.6	19677.4	49.0	48.9	49.1	48.0
τ=-0.2	Bias	.004	.022	.018	.002	.004	.022	.018	.002	.005	.015	.002	.004	.005	.015	.002	.004
	RMSE	.158	.134	.143	.154	.158	.134	.143	.154	.099	.110	.095	.108	.099	.110	.095	.108
	Coverage	.940	.970	.950	.930	.940	.970	.950	.930	.950	.930	.950	.910	.950	.930	.950	.910
	Time	2562.3	2314.1	2315.7	2576.4	37.2	37.5	37.4	36.6	17579.9	15779.5	15744.3	17671.9	48.9	48.6	49.0	48.0
τ=0.2	Bias	$-.015$	.012	$-.003$	.019	$-.015$	.012	$-.003$	.019	$-.006$	.000	.000	.006	$-.006$	.000	.000	.006
	RMSE	.165	.133	.141	.145	.165	.133	.141	.145	.094	.106	.104	.114	.094	.106	.104	.114
	Coverage	.870	.960	.950	.970	.870	.960	.950	.970	.960	.940	.920	.920	.960	.940	.920	.920
	Time	2560.7	2328.0	2313.5	2561.8	37.4	36.8	37.1	36.9	17602.7	15790.3	15729.3	17749.5	49.7	48.1	48.8	47.8
τ=2	Bias	$-.042$	$-.047$	$-.035$	$-.055$	$-.042$	$-.047$	$-.035$	$-.055$	$-.016$	$-.013$	$-.016$	$-.014$	$-.016$	$-.013$	$-.016$	$-.014$
	RMSE	.141	.155	.163	.164	.141	.155	.163	.164	.101	.103	.106	.102	.101	.103	.106	.102
	Coverage	.980	.960	.940	.960	.980	.960	.940	.960	.940	.980	.940	.970	.940	.980	.940	.970
	Time	2758.7	2504.3	2518.9	2769.1	37.5	36.9	36.7	36.3	19622.7	17732.9	17730.0	19463.7	49.0	48.4	48.2	47.8

5 An Empirical Illustration

In this section, we illustrate the computation time advantage of the matrix-vector products method using an empirical application. To this end, we use an example from Pace and Barry (1997) on the US presidential election in 1980. The dataset contains variables on the election results and county characteristics for 3107 US counties. In our model, the outcome variable is the (logged) proportion of voting age population that voted in the election (Y = Vote). The explanatory variables include the log percentage of population with a twelfth grade or higher education (X_1 = Educ), the log percentage of population with homeownership (X_2 = Homeowners), and the log per capita income (X_3 = Income). We consider the following MESS(1,1) specification

$$e^{\alpha W}Y = \beta_0 l_n + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u, \quad e^{\tau W}u = \epsilon. \tag{5.1}$$

The spatial weights matrix W is the contiguity based weights matrix constructed using the latitude and longitude of the counties. We estimate this specification using the QML, GMM and Bayesian methods. For the Bayesian method, the same priors as those in the MC simulations are used. We use the expm and mvp methods to compute the estimation results and record the corresponding computation times. In the case of mvp method, the truncation order q is set to 15.

The parameter estimation results are summarized in Table 6. While the mvp method provides the same parameter estimates and standard errors (in parenthesis) as those obtained from the expm method, the computation time is significantly smaller for the mvp method. In the QML case, the computation time is 1072.4 seconds for the expm method, but only 6.0 seconds for the mvp method. In the GMM case, the computation time takes 4805.8. seconds for the expm method, and 21.1 seconds for the mvp method. For the Bayesian estimation, the number of draws is set to 1500 and the first 500 draws are discarded as burn-ins. The results show that the difference in terms of computation time is the biggest for the Bayesian estimation case, costing 47742.3. seconds (13.26 hours) for the expm method, and 11.4 seconds for the mvp method. Overall, these results show that the matrix-vector products method is not only useful in Monte Carlo simulations, where resamples are drawn for hundreds or thousands of times, but also useful in empirical applications with large sample sizes.

To investigate the impact measures, we also compute the average direct effects, average total effects and average indirect effects discussed in section 2.5, the measures of dispersion and computing time. The results are summarized in Table 7. To compute the dispersion of the impact measure we use the delta method in section 2.5 for QML and GMM methods and use the standard deviation of posterior draws of the impact measure for the Bayesian method. We can see that the corresponding impact measures are similar for different computation methods, i.e., QMLE, GMME and Bayesian method. However, the mvp method takes much less time than the expm method. For example, the average direct effect for *Educ* is 0.320 using QML method, 0.305 using GMM method and 0.320 for Bayesian method. However, the computation time for the expm method are 1088.2, 4822.3 and 71584.0 seconds for QML, GMM and Bayesian method respectively and 22.0, 37.6 and 34491.0

seconds for the mvp method respectively[11].

Table 6: The parameter estimates of the presidential election voting example

	QML		GMM		Bayesian	
	expm	mvp	expm	mvp	expm	mvp
α	-0.350^{***}	-0.350^{***}	-0.423^{***}	-0.423^{***}	-0.337^{***}	-0.337^{***}
	(0.045)	(0.045)	(0.045)	(0.045)	(0.041)	(0.041)
τ	-0.443^{***}	-0.443^{***}	-0.374^{***}	-0.374^{***}	-0.458^{***}	-0.458^{***}
	(0.055)	(0.055)	(0.055)	(0.055)	(0.050)	(0.050)
Educ	0.316^{***}	0.316^{***}	0.300^{***}	0.300^{***}	0.317^{***}	0.317^{***}
	(0.021)	(0.021)	(0.020)	(0.020)	(0.020)	(0.020)
Homeowners	0.572^{***}	0.572^{***}	0.571^{***}	0.571^{***}	0.572^{***}	0.572^{***}
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Income	-0.154^{***}	-0.154^{***}	-0.144^{***}	-0.144^{***}	-0.155^{***}	-0.155^{***}
	(0.021)	(0.021)	(0.020)	(0.020)	(0.021)	(0.021)
Cons.	0.738^{***}	0.738^{***}	0.732^{***}	0.732^{***}	0.734^{***}	0.734^{***}
	(0.052)	(0.052)	(0.051)	(0.051)	(0.054)	(0.054)
Time	1072.4	6.0	4805.8	21.1	47742.3	11.4

[11] Note that the impact measures consists of traces of $n \times n$ matrices. For example, the average direct effect equals $\frac{1}{n}\mathrm{tr}(e^{-\widehat{\alpha}W_n}\widehat{\beta}_k)$ for $k = 1, 2$ and 3, so we need to compute the trace of an $n \times n$ matrix for each draw in the bayesian method. This is why the time to compute the impact measures is long using Bayesian method. However the time to sample the draws for the parameters are much faster as shown in Table 6.

Table 7: The impact measures of the presidential election voting example

	QML		GMM		Bayesian	
	expm	mvp	expm	mvp	expm	mvp
			Average direct effects			
Educ	0.320***	0.320***	0.305***	0.305***	0.320***	0.320***
	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
Homeowners	0.578***	0.578***	0.580***	0.580***	0.578***	0.578***
	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)	(0.016)
Income	−0.156***	−0.156***	−0.147***	−0.147***	−0.156***	−0.156***
	(0.021)	(0.021)	(0.020)	(0.020)	(0.021)	(0.021)
			Average total effects			
Educ	0.449***	0.449***	0.458***	0.458***	0.444***	0.444***
	(0.027)	(0.027)	(0.027)	(0.027)	(0.029)	(0.029)
Homeowners	0.812***	0.812***	0.872***	0.872***	0.802***	0.802***
	(0.043)	(0.043)	(0.044)	(0.044)	(0.039)	(0.039)
Income	−0.219***	−0.219***	−0.220***	−0.220***	−0.217***	−0.217***
	(0.028)	(0.028)	(0.030)	(0.030)	(0.029)	(0.029)
			Average indirect effects			
Educ	0.129***	0.129***	0.153***	0.153***	0.124***	0.124***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.018)	(0.018)
Homeowners	0.234***	0.234***	0.292***	0.292***	0.224***	0.224***
	(0.036)	(0.036)	(0.038)	(0.038)	(0.032)	(0.032)
Income	−0.063***	−0.063***	−0.074***	−0.074***	−0.060***	−0.060***
	(0.011)	(0.011)	(0.012)	(0.012)	(0.011)	(0.011)
Time	1088.2	22.0	4822.3	37.6	71584.0	24491.0

6 Conclusion

The MESS-type models provide a class of alternatives to SAR-type models with attractive properties. However, the estimation of these models require the computation of the matrix exponential terms in each iteration of a numerical optimization scheme, which can be computationally costly. In this paper, for the calculation of the matrix exponential terms, we proposed a matrix-vector products method based on the truncation of Taylor series expansion of matrix exponential terms. Because the estimation of MESS-type models requires the computation of the matrix exponential terms as vectors, rather than the matrix exponential terms in isolation, our approach provides an efficient alternative to the default method available in several popular statistical software. The results from our extensive simulation study and empirical illustration confirmed the computational time gains for three estimation methods for MESS-type models (i.e., the QML, GMM and Bayesian methods) using the matrix-vector products method.

References

Al-Mohy, Awad H. and Nicholas J. Higham (2010). "A New Scaling and Squaring Algorithm for the Matrix Exponential". In: *SIAM Journal on Matrix Analysis and Applications* 31.3, pp. 970–989.

Anselin, Luc (1988). *Spatial econometrics: Methods and Models.* New York: Springer.

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

Bader, Philipp, Sergio Blanes, and Fernando Casas (2019). "Computing the matrix exponential with an optimized Taylor polynomial approximation". In: *Mathematics* 7.12, p. 1174.

Chiu, Tom Y. M., Tom Leonard, and Kam-Wah Tsui (1996). "The Matrix-Logarithmic Covariance Model". In: *Journal of the American Statistical Association* 91.433, pp. 198–210.

Cliff, A. D. and J. K. Ord (1969). "The Problem of Spatial Autocorrelation". In: ed. by A. J. Scott. Vol. In London Papers in Regional Science 1. Studies in Regional Science. London: Pion, pp. 25–55.

— (1973). *Spatial Autocorrelation.* London: Pion.

Debarsy, Nicolas, Fei Jin, and Lung fei Lee (2015). "Large sample properties of the matrix exponential spatial specification with an application to FDI". In: *Journal of Econometrics* 188.1, pp. 1–21.

Elhorst, J. Paul (2014). *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels.* SpringerBriefs in Regional Science. Springer Berlin Heidelberg.

Gallopoulos, E. and Y. Saad (1992). "Efficient Solution of Parabolic Equations by Krylov Approximation Methods". In: *SIAM Journal on Scientific and Statistical Computing* 13.5, pp. 1236–1264.

Higham, Nicholas J (2005). "The scaling and squaring method for the matrix exponential revisited". In: *SIAM Journal on Matrix Analysis and Applications* 26.4, pp. 1179–1193.

Hochbruck, Marlis and Christian Lubich (1997). "On Krylov Subspace Approximations to the Matrix Exponential Operator". In: *SIAM Journal on Numerical Analysis* 34.5, pp. 1911–1925.

Horn, R.A. and C.R. Johnson (2012). *Matrix Analysis.* Matrix Analysis. Cambridge University Press. ISBN: 9780521839402.

Kelejian, Harry H. and Ingmar R. Prucha (2010). "Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances". In: *Journal of Econometrics* 157.1, pp. 53 –67.

Leonard, Tom and John S. J. Hsu (1992). "Bayesian Inference for a Covariance Matrix". In: *The Annals of Statistics* 20.4, pp. 1669 –1696.

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

— (2009). *Introduction to Spatial Econometrics.* Florida: CRC Press.

Moler, Cleve and Charles Van Loan (1978). "Nineteen dubious ways to compute the exponential of a matrix". In: *SIAM review* 20.4, pp. 801–836.

Moler, Cleve and Charles Van Loan (2003). "Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later". In: *SIAM review* 45.1, pp. 3–49.

Pace, R. Kelley and Ronald Barry (1997). "Quick Computation of Spatial Autoregressive Estimators". In: *Geographical Analysis* 29.3, pp. 232–247.

Saad, Y. (1992). "Analysis of Some Krylov Subspace Approximations to the Matrix Exponential Operator". In: *SIAM Journal on Numerical Analysis* 29.1, pp. 209–228.

Sidje, Roger B (1998). "Expokit: A software package for computing matrix exponentials". In: *ACM Transactions on Mathematical Software (TOMS)* 24.1, pp. 130–156.

Taşpınar, Süleyman, Osman Doğan, and Wim P. M. Vijverberg (2018). "GMM inference in spatial autoregressive models". In: *Econometric Reviews* 37.9, pp. 931–954.

Whittle, P. (1954). "On Stationary Processes in the Plane". In: *Biometrika* 41.3/4, pp. 434–449.

Appendix

A Additional Simulation Results

In this appendix, we provide some additional simulation results on the performance of both methods. Tables A.1 and A.2 include the simulation results for the indirect and total effects of X_1 and X_2 for the QMLE case. Tables A.3–A.6 provide additional results for the GMME. The remaining tables, Tables A.7–A.10, include additional simulation results for the Bayesian estimator. Overall, the simulation results in these tables attest that the mvp method is computationally more efficient than the expm method.

Table A.1: The QMLE results of average indirect effects for X_1 and X_2

		n=169								n=361							
		expm				mvp				expm				mvp			
		α				α				α				α			
		−2	−0.2	0.2	2	−2	−0.2	0.2	2	−2	−0.2	0.2	2	−2	−0.2	0.2	2
Results for X_1																	
τ=−2	Bias	−.006	−.005	.002	−.001	−.006	−.005	.002	−.001	.018	.000	.000	.001	.018	.000	.000	.001
	RMSE	.547	.079	.056	.107	.547	.079	.056	.107	.406	.054	.042	.071	.406	.054	.042	.071
	Coverage	.931	.943	.935	.945	.931	.943	.935	.945	.942	.951	.955	.943	.942	.951	.955	.943
	Time	613.8	614.1	597.3	626.7	15.5	15.0	14.6	15.3	4217.8	4306.7	4188.5	4277.7	55.7	58.2	57.3	55.3
τ=−0.2	Bias	.029	−.005	.000	−.008	.029	−.005	.000	−.008	.007	−.002	.002	−.002	.007	−.002	.002	−.002
	RMSE	.875	.126	.098	.152	.875	.126	.098	.152	.615	.090	.065	.101	.615	.090	.065	.101
	Coverage	.932	.939	.941	.940	.932	.939	.941	.940	.949	.945	.945	.938	.949	.945	.945	.938
	Time	643.9	559.0	578.3	668.2	17.6	15.6	16.0	18.0	4196.7	3587.0	3709.6	4271.9	62.5	58.4	59.9	62.9
τ=0.2	Bias	−.012	−.002	−.002	−.002	−.012	−.002	−.002	−.002	.007	.004	.000	.000	.007	.004	.000	.000
	RMSE	.853	.120	.092	.166	.853	.120	.092	.166	.603	.092	.073	.104	.603	.092	.073	.104
	Coverage	.946	.942	.945	.939	.946	.942	.945	.939	.947	.953	.929	.938	.947	.953	.929	.938
	Time	598.0	540.6	555.6	631.3	16.5	15.2	15.5	17.2	3885.7	3494.3	3538.3	4055.5	59.1	57.4	57.9	60.6
τ=2	Bias	.031	.001	−.001	.001	.031	.001	−.001	.001	.012	.001	.001	.001	.012	.001	.001	.001
	RMSE	.420	.049	.048	.083	.420	.049	.048	.083	.287	.041	.037	.058	.287	.041	.037	.058
	Coverage	.918	.936	.943	.942	.918	.936	.943	.942	.925	.943	.944	.942	.925	.943	.944	.942
	Time	639.1	657.8	618.8	650.9	15.4	15.7	14.8	15.6	4403.5	4555.0	4312.5	4561.4	55.6	60.3	58.4	56.8
Results for X_2																	
τ=−2	Bias	−.008	−.002	.001	−.006	−.008	−.002	.001	−.006	.012	.000	.000	.000	.012	.000	.000	.000
	RMSE	.393	.042	.029	.101	.393	.042	.029	.101	.284	.029	.021	.068	.284	.029	.021	.068
	Coverage	.937	.931	.943	.928	.937	.931	.943	.928	.942	.954	.968	.896	.942	.954	.968	.896
	Time	613.8	614.1	597.3	626.7	15.5	15.0	14.6	15.3	4217.8	4306.7	4188.5	4277.7	55.6	58.2	57.3	55.4
τ=−0.2	Bias	.009	−.003	.001	−.003	.009	−.003	.001	−.003	.020	−.001	.001	.000	.020	−.001	.001	.000
	RMSE	.615	.063	.048	.118	.615	.063	.048	.118	.415	.047	.033	.083	.415	.047	.033	.083
	Coverage	.922	.946	.954	.967	.922	.946	.954	.967	.959	.938	.949	.951	.959	.938	.949	.951
	Time	643.9	559.0	578.3	668.2	17.6	15.6	16.0	18.0	4196.7	3587.0	3709.6	4271.9	62.4	58.3	59.9	62.9
τ=0.2	Bias	−.003	.000	.000	.002	−.003	.000	.000	.002	.002	.002	.000	−.003	.002	.002	.000	−.003
	RMSE	.543	.063	.048	.147	.543	.063	.048	.147	.401	.047	.037	.084	.401	.047	.037	.084
	Coverage	.964	.933	.947	.917	.964	.933	.947	.917	.953	.958	.935	.952	.953	.958	.935	.952
	Time	598.0	540.6	555.6	631.3	16.5	15.2	15.5	17.1	3885.7	3494.2	3538.4	4055.5	59.1	57.4	57.9	60.6
τ=2	Bias	.012	.001	.000	.003	.012	.001	.000	.003	−.001	.001	.000	.001	−.001	.001	.000	.001
	RMSE	.301	.024	.021	.058	.301	.024	.021	.058	.171	.021	.018	.042	.171	.021	.018	.042
	Coverage	.907	.944	.966	.981	.907	.944	.966	.981	.967	.917	.947	.958	.967	.917	.947	.958
	Time	639.1	657.7	618.8	650.9	15.4	15.7	14.8	15.6	4403.5	4555.1	4312.5	4561.4	55.5	60.4	58.4	56.8

Table A.2: The QMLE results of average total effects for X_1 and X_2

		n=169								n=361							
		expm				mvp				expm				mvp			
		α				α				α				α			
		−2	−0.2	0.2	2	−2	−0.2	0.2	2	−2	−0.2	0.2	2	−2	−0.2	0.2	2
Results for X_1																	
τ=−2	Bias	−.003	−.002	−.002	−.001	−.003	−.002	−.002	−.001	.021	.004	.000	.000	.021	.004	.000	.000
	RMSE	.643	.103	.072	.013	.643	.103	.072	.013	.477	.073	.047	.008	.477	.073	.047	.008
	Coverage	.938	.948	.936	.940	.938	.948	.936	.940	.941	.954	.945	.944	.941	.954	.945	.944
	Time	612.0	612.7	595.9	625.0	13.8	13.5	13.2	13.6	4205.5	4297.0	4178.6	4265.6	43.4	48.5	47.5	43.3
τ=−0.2	Bias	.034	−.002	−.001	−.001	.034	−.002	−.001	−.001	.006	−.002	−.002	.000	.006	−.002	−.002	.000
	RMSE	1.021	.166	.107	.021	1.021	.166	.107	.021	.712	.108	.075	.013	.712	.108	.075	.013
	Coverage	.937	.934	.940	.934	.937	.934	.940	.934	.953	.950	.931	.952	.953	.950	.931	.952
	Time	642.2	557.5	576.8	666.5	15.9	14.2	14.5	16.2	4184.3	3577.2	3699.7	4259.5	50.0	48.5	50.0	50.6
τ=0.2	Bias	−.015	−.002	−.003	.000	−.015	−.002	−.003	.000	.010	.004	.000	.001	.010	.004	.000	.001
	RMSE	1.003	.142	.108	.020	1.003	.142	.108	.020	.700	.109	.079	.013	.700	.109	.079	.013
	Coverage	.945	.938	.936	.926	.945	.938	.936	.926	.944	.953	.926	.937	.944	.953	.926	.937
	Time	596.2	539.1	554.2	629.6	14.8	13.8	14.1	15.4	3873.4	3484.4	3528.5	4043.1	46.8	47.5	48.0	48.3
τ=2	Bias	.036	.001	−.002	.000	.036	.001	−.002	.000	.015	−.001	−.001	.000	.015	−.001	−.001	.000
	RMSE	.496	.061	.053	.010	.496	.061	.053	.010	.338	.046	.042	.004	.338	.046	.042	.004
	Coverage	.921	.942	.928	.931	.921	.942	.928	.931	.924	.945	.930	.940	.924	.945	.930	.940
	Time	637.3	656.3	617.3	649.2	13.7	14.3	13.3	13.9	4391.6	4545.3	4302.6	4549.4	43.6	50.6	48.6	44.8
Results for X_2																	
τ=−2	Bias	−.009	−.003	.000	.000	−.009	−.003	.000	.000	.014	−.001	−.002	.000	.014	−.001	−.002	.000
	RMSE	.484	.076	.049	.009	.484	.076	.049	.009	.349	.055	.037	.006	.349	.055	.037	.006
	Coverage	.951	.947	.934	.938	.951	.947	.934	.938	.942	.948	.942	.942	.942	.948	.942	.942
	Time	612.0	612.7	595.9	624.9	13.7	13.5	13.1	13.6	4205.5	4297.0	4178.6	4265.6	43.4	48.5	47.5	43.2
τ=−0.2	Bias	.010	−.003	−.001	.000	.010	−.003	−.001	.000	.024	−.003	−.002	.000	.024	−.003	−.002	.000
	RMSE	.752	.110	.081	.014	.752	.110	.081	.014	.504	.079	.051	.009	.504	.079	.051	.009
	Coverage	.930	.936	.944	.922	.930	.936	.944	.922	.945	.946	.935	.948	.945	.946	.935	.948
	Time	642.2	557.5	576.8	666.5	15.9	14.2	14.5	16.2	4184.3	3577.1	3699.7	4259.5	50.1	48.5	50.0	50.6
τ=0.2	Bias	−.003	.002	−.005	.000	−.003	.002	−.005	.000	.003	.004	.001	.000	.003	.004	.001	.000
	RMSE	.667	.118	.077	.015	.667	.118	.077	.015	.487	.076	.056	.009	.487	.076	.056	.009
	Coverage	.941	.935	.937	.936	.941	.935	.937	.936	.959	.951	.937	.946	.959	.951	.937	.946
	Time	596.2	539.1	554.1	629.6	14.8	13.8	14.0	15.4	3873.4	3484.4	3528.4	4043.1	46.8	47.5	48.0	48.3
τ=2	Bias	.014	.002	−.001	.000	.014	.002	−.001	.000	−.002	.003	.001	.000	−.002	.003	.001	.000
	RMSE	.367	.043	.049	.007	.367	.043	.049	.007	.208	.037	.027	.004	.208	.037	.027	.004
	Coverage	.922	.936	.940	.930	.922	.936	.940	.930	.946	.948	.949	.939	.946	.948	.949	.939
	Time	637.4	656.3	617.3	649.2	13.7	14.3	13.3	13.9	4391.6	4545.3	4302.6	4549.4	43.6	50.6	48.6	44.8

Table A.3: The GMME results for $\widehat{\beta}_1$ and $\widehat{\beta}_2$

		n=169								n=361							
		expm				mvp				expm				mvp			
		α				α				α				α			
		-2	-0.2	0.2	2	-2	-0.2	0.2	2	-2	-0.2	0.2	2	-2	-0.2	0.2	2
Results for $\widehat{\beta}_1$																	
$\tau=-2$	Bias	.002	.003	$-.005$	$-.003$	.002	.003	$-.005$	$-.003$	.000	.003	.000	$-.001$	.000	.003	.000	$-.001$
	RMSE	.060	.061	.057	.059	.060	.061	.057	.059	.040	.039	.037	.035	.040	.039	.037	.035
	Coverage	.930	.931	.935	.919	.930	.931	.935	.919	.950	.937	.939	.945	.950	.937	.939	.945
	Time	1407.9	1298.5	1257.3	1270.5	60.6	57.0	56.2	54.9	8663.3	8788.2	8576.1	8797.2	219.6	220.6	219.6	219.7
$\tau=-0.2$	Bias	$-.001$	$-.001$	$-.004$	$-.004$	$-.001$	$-.001$	$-.004$	$-.004$	$-.003$	$-.002$	$-.006$	.000	$-.003$	$-.002$	$-.006$	.000
	RMSE	.081	.078	.076	.086	.081	.078	.076	.086	.052	.053	.055	.053	.052	.053	.055	.053
	Coverage	.934	.931	.942	.916	.934	.931	.942	.916	.949	.936	.931	.938	.949	.936	.931	.938
	Time	1512.6	1413.2	1378.5	1504.4	60.8	59.7	58.8	60.5	9794.5	8872.9	8787.4	9936.7	207.6	203.1	202.2	208.4
$\tau=0.2$	Bias	$-.004$	$-.004$	$-.005$	$-.006$	$-.004$	$-.004$	$-.005$	$-.006$	.000	$-.002$	$-.002$	$-.001$	.000	$-.002$	$-.002$	$-.001$
	RMSE	.086	.077	.082	.082	.086	.077	.082	.082	.054	.054	.054	.057	.054	.054	.054	.057
	Coverage	.925	.926	.924	.934	.925	.926	.924	.934	.950	.946	.934	.935	.950	.946	.934	.935
	Time	1409.1	1312.1	1300.7	1409.1	58.5	57.5	57.2	58.6	9341.3	8504.4	8409.4	9463.1	205.9	201.4	200.1	205.8
$\tau=2$	Bias	.001	.000	$-.002$	$-.001$	.001	.000	$-.002$	$-.001$	.001	$-.002$	$-.002$	.000	.001	$-.002$	$-.002$	.000
	RMSE	.037	.042	.038	.047	.037	.042	.038	.047	.029	.026	.030	.027	.029	.026	.030	.027
	Coverage	.909	.910	.932	.930	.909	.910	.932	.930	.940	.930	.938	.935	.940	.930	.938	.935
	Time	1420.2	1335.1	1340.9	1343.2	58.4	58.2	58.4	56.4	9321.8	9157.7	8956.1	9167.3	223.0	224.4	222.9	221.0
Results for $\widehat{\beta}_2$																	
$\tau=-2$	Bias	.000	.000	$-.001$	.002	.000	.000	$-.001$	.002	.001	$-.001$	$-.003$	$-.001$	.001	$-.001$	$-.003$	$-.001$
	RMSE	.058	.052	.050	.062	.058	.052	.050	.062	.039	.038	.038	.042	.039	.038	.038	.042
	Coverage	.924	.947	.904	.926	.924	.947	.904	.926	.941	.950	.940	.930	.941	.950	.940	.930
	Time	1407.9	1298.5	1257.3	1270.5	60.6	57.0	56.2	54.9	8663.3	8788.2	8576.1	8797.2	219.6	220.6	219.6	219.7
$\tau=-0.2$	Bias	$-.002$	$-.002$	$-.003$	$-.001$	$-.002$	$-.002$	$-.003$	$-.001$	.001	$-.003$	$-.004$	$-.001$	.001	$-.003$	$-.004$	$-.001$
	RMSE	.083	.078	.078	.078	.083	.078	.078	.078	.055	.054	.051	.052	.055	.054	.051	.052
	Coverage	.915	.921	.940	.932	.915	.921	.940	.932	.935	.943	.937	.932	.935	.943	.937	.932
	Time	1512.6	1413.2	1378.5	1504.4	60.8	59.7	58.8	60.5	9794.5	8872.9	8787.4	9936.7	207.6	203.1	202.2	208.4
$\tau=0.2$	Bias	.000	.001	$-.007$	$-.006$	.000	.001	$-.007$	$-.006$	$-.001$	.001	.001	.001	$-.001$	.001	.001	.001
	RMSE	.079	.089	.083	.094	.079	.089	.083	.094	.053	.054	.053	.052	.053	.054	.053	.052
	Coverage	.929	.914	.920	.919	.929	.914	.920	.919	.948	.930	.940	.938	.948	.930	.940	.938
	Time	1409.1	1312.1	1300.7	1409.1	58.5	57.5	57.2	58.6	9341.3	8504.4	8409.4	9463.1	205.9	201.4	200.1	205.8
$\tau=2$	Bias	.000	.000	$-.001$	$-.002$	.000	.000	$-.001$	$-.002$	$-.001$	.002	.000	.000	$-.001$	.002	.000	.000
	RMSE	.035	.036	.044	.038	.035	.036	.044	.038	.022	.026	.024	.028	.022	.026	.024	.028
	Coverage	.916	.925	.924	.915	.916	.925	.924	.915	.947	.938	.938	.926	.947	.938	.938	.926
	Time	1420.2	1335.1	1340.9	1343.2	58.4	58.2	58.4	56.4	9321.8	9157.7	8956.1	9167.3	223.0	224.4	222.9	221.0

Table A.4: The GMME results of average direct effects for X_1 and X_2

		n=169								n=361							
		expm				mvp				expm				mvp			
		α				α				α				α			
		−2	−0.2	0.2	2	−2	−0.2	0.2	2	−2	−0.2	0.2	2	−2	−0.2	0.2	2
Results for X_1																	
τ=−2	Bias	.002	.003	−.005	−.004	.002	.003	−.005	−.004	.003	.003	.000	−.003	.003	.003	.000	−.003
	RMSE	.109	.061	.057	.113	.109	.061	.057	.113	.078	.039	.037	.071	.078	.039	.037	.071
	Coverage	.928	.935	.934	.918	.928	.935	.934	.918	.940	.934	.935	.944	.940	.934	.935	.944
	Time	1409.8	1299.9	1258.7	1272.2	62.5	58.4	57.7	56.7	8675.5	8798.0	8585.9	8809.3	231.8	230.4	229.4	231.8
τ=−0.2	Bias	.003	.000	−.003	−.001	.003	.000	−.003	−.001	−.002	−.001	−.005	.000	−.002	−.001	−.005	.000
	RMSE	.162	.079	.077	.157	.162	.079	.077	.157	.107	.053	.055	.103	.107	.053	.055	.103
	Coverage	.932	.928	.941	.934	.932	.928	.941	.934	.949	.935	.930	.928	.949	.935	.930	.928
	Time	1514.4	1414.7	1380.0	1506.1	62.6	61.1	60.2	62.3	9807.0	8882.8	8797.4	9949.1	220.1	213.0	212.1	220.8
τ=0.2	Bias	−.005	−.003	−.004	−.008	−.005	−.003	−.004	−.008	.001	−.002	−.001	−.003	.001	−.002	−.001	−.003
	RMSE	.168	.076	.083	.168	.168	.076	.083	.168	.108	.053	.055	.107	.108	.053	.055	.107
	Coverage	.933	.925	.926	.925	.933	.925	.926	.925	.946	.943	.933	.932	.946	.943	.933	.932
	Time	1410.8	1313.5	1302.2	1410.8	60.2	58.9	58.7	60.3	9353.7	8514.4	8419.4	9475.4	218.2	211.3	210.1	218.1
τ=2	Bias	.004	.000	−.001	−.004	.004	.000	−.001	−.004	.003	−.002	−.002	−.002	.003	−.002	−.002	−.002
	RMSE	.083	.041	.039	.087	.083	.041	.039	.087	.056	.026	.030	.059	.056	.026	.030	.059
	Coverage	.906	.910	.932	.931	.906	.910	.932	.931	.927	.933	.938	.932	.927	.933	.938	.932
	Time	1421.9	1336.5	1342.3	1344.9	60.1	59.6	59.9	58.1	9333.7	9167.5	8966.0	9179.2	235.0	234.2	232.7	232.8
Results for X_2																	
τ=−2	Bias	−.001	.000	−.001	.005	−.001	.000	−.001	.005	.003	−.001	−.003	−.001	.003	−.001	−.003	−.001
	RMSE	.100	.053	.050	.111	.100	.053	.050	.111	.069	.038	.038	.073	.069	.038	.038	.073
	Coverage	.923	.946	.908	.930	.923	.946	.908	.930	.932	.947	.943	.930	.932	.947	.943	.930
	Time	1409.9	1299.9	1258.7	1272.2	62.5	58.4	57.6	56.7	8675.5	8798.1	8586.0	8809.2	231.9	230.4	229.4	231.8
τ=−0.2	Bias	−.001	−.002	−.003	.000	−.001	−.002	−.003	.000	.003	−.003	−.004	−.002	.003	−.003	−.004	−.002
	RMSE	.149	.079	.078	.130	.149	.079	.078	.130	.097	.054	.052	.090	.097	.054	.052	.090
	Coverage	.917	.922	.941	.934	.917	.922	.941	.934	.945	.943	.936	.927	.945	.943	.936	.927
	Time	1514.4	1414.7	1380.0	1506.1	62.6	61.1	60.2	62.3	9807.0	8882.8	8797.4	9949.1	220.1	213.1	212.1	220.8
τ=0.2	Bias	.000	.002	−.006	−.009	.000	.002	−.006	−.009	.000	.001	.001	.001	.000	.001	.001	.001
	RMSE	.136	.089	.083	.163	.136	.089	.083	.163	.094	.054	.053	.090	.094	.054	.053	.090
	Coverage	.931	.913	.923	.924	.931	.913	.923	.924	.952	.929	.940	.942	.952	.929	.940	.942
	Time	1410.8	1313.5	1302.2	1410.8	60.2	58.9	58.7	60.3	9353.7	8514.3	8419.3	9475.4	218.2	211.4	210.1	218.1
τ=2	Bias	.002	.000	−.001	−.004	.002	.000	−.001	−.004	−.001	.002	.000	−.001	−.001	.002	.000	−.001
	RMSE	.070	.035	.043	.064	.070	.035	.043	.064	.039	.026	.024	.046	.039	.026	.024	.046
	Coverage	.919	.924	.925	.916	.919	.924	.925	.916	.943	.939	.940	.926	.943	.939	.940	.926
	Time	1421.9	1336.5	1342.3	1344.9	60.1	59.6	59.8	58.1	9333.7	9167.6	8966.0	9179.2	235.0	234.3	232.7	232.9

Table A.5: The GMME results of average indirect effects for X_1 and X_2

		n=169								n=361							
		expm				mvp				expm				mvp			
		τ				τ				τ				τ			
		−2	−0.2	0.2	2	−2	−0.2	0.2	2	−2	−0.2	0.2	2	−2	−0.2	0.2	2
Results for X_1																	
tau=−2	Bias	.000	−.002	.004	.003	.000	−.002	.004	.003	.026	.002	.001	.003	.026	.002	.001	.003
	RMSE	.567	.081	.059	.112	.567	.081	.059	.112	.416	.056	.043	.071	.416	.056	.043	.071
	Coverage	.929	.931	.918	.916	.929	.931	.918	.916	.940	.949	.947	.941	.940	.949	.947	.941
	Time	1411.7	1301.4	1260.1	1273.9	64.6	59.8	59.1	58.4	8687.6	8807.8	8595.7	8821.3	244.0	240.1	239.2	243.8
tau=−0.2	Bias	.031	−.002	.002	.001	.031	−.002	.002	.001	.015	−.001	.004	.001	.015	−.001	.004	.001
	RMSE	.884	.131	.102	.157	.884	.131	.102	.157	.628	.091	.066	.103	.628	.091	.066	.103
	Coverage	.926	.924	.933	.937	.926	.924	.933	.937	.940	.943	.939	.930	.940	.943	.939	.930
	Time	1516.2	1416.1	1381.4	1507.8	64.3	62.6	61.7	64.0	9819.3	8892.7	8807.3	9961.4	232.5	222.9	222.0	233.0
tau=0.2	Bias	−.009	−.002	.000	.008	−.009	−.002	.000	.008	.010	.006	.001	.004	.010	.006	.001	.004
	RMSE	.883	.125	.094	.171	.883	.125	.094	.171	.610	.094	.074	.107	.610	.094	.074	.107
	Coverage	.930	.926	.936	.924	.930	.926	.936	.924	.941	.946	.924	.936	.941	.946	.924	.936
	Time	1412.5	1315.0	1303.6	1412.6	61.9	60.3	60.1	62.0	9365.9	8524.2	8429.3	9487.7	230.4	221.2	220.0	230.4
tau=2	Bias	.030	.002	.001	.005	.030	.002	.001	.005	.015	.002	.002	.002	.015	.002	.002	.002
	RMSE	.438	.052	.052	.086	.438	.052	.052	.086	.292	.041	.038	.059	.292	.041	.038	.059
	Coverage	.907	.930	.930	.929	.907	.930	.930	.929	.919	.936	.939	.932	.919	.936	.939	.932
	Time	1423.6	1337.9	1343.7	1346.6	61.7	61.1	61.3	59.8	9345.6	9177.3	8975.7	9191.0	246.9	244.0	242.4	244.7
Results for X_2																	
tau=−2	Bias	−.007	−.001	.002	−.005	−.007	−.001	.002	−.005	.017	.000	.001	.002	.017	.000	.001	.002
	RMSE	.411	.043	.031	.105	.411	.043	.031	.105	.290	.029	.022	.069	.290	.029	.022	.069
	Coverage	.911	.915	.923	.899	.911	.915	.923	.899	.932	.951	.961	.883	.932	.951	.961	.883
	Time	1411.7	1301.4	1260.1	1273.9	64.6	59.8	59.1	58.4	8687.5	8807.8	8595.7	8821.3	243.9	240.1	239.2	243.9
tau=−0.2	Bias	.010	−.002	.002	−.001	.010	−.002	.002	−.001	.022	−.001	.002	.003	.022	−.001	.002	.003
	RMSE	.632	.066	.050	.122	.632	.066	.050	.122	.425	.047	.034	.085	.425	.047	.034	.085
	Coverage	.907	.931	.939	.952	.907	.931	.939	.952	.945	.932	.936	.944	.945	.932	.936	.944
	Time	1516.1	1416.1	1381.4	1507.8	64.4	62.6	61.7	64.0	9819.4	8892.7	8807.3	9961.4	232.5	222.9	222.0	233.0
tau=0.2	Bias	.002	.000	.001	.008	.002	.000	.001	.008	.002	.003	.000	−.001	.002	.003	.000	−.001
	RMSE	.564	.066	.050	.154	.564	.066	.050	.154	.406	.048	.037	.085	.406	.048	.037	.085
	Coverage	.945	.915	.935	.887	.945	.915	.935	.887	.946	.953	.931	.945	.946	.953	.931	.945
	Time	1412.6	1315.0	1303.6	1412.6	61.9	60.3	60.1	62.0	9365.9	8524.3	8429.3	9487.7	230.4	221.3	220.0	230.4
tau=2	Bias	.017	.001	.001	.004	.017	.001	.001	.004	−.001	.002	.001	.001	−.001	.002	.001	.001
	RMSE	.310	.026	.023	.061	.310	.026	.023	.061	.174	.022	.019	.043	.174	.022	.019	.043
	Coverage	.899	.935	.961	.970	.899	.935	.961	.970	.961	.914	.941	.959	.961	.914	.941	.959
	Time	1423.6	1337.9	1343.7	1346.6	61.7	61.1	61.3	59.8	9345.6	9177.3	8975.7	9191.0	247.0	244.0	242.4	244.6

Table A.6: The GMME results of average total effects for X_1 and X_2

		n=169								n=361							
		expm				mvp				expm				mvp			
		α				α				α				α			
		-2	-0.2	0.2	2	-2	-0.2	0.2	2	-2	-0.2	0.2	2	-2	-0.2	0.2	2
Results for X_1																	
τ=-2	Bias	.002	.000	$-.001$	.000	.002	.000	$-.001$	.000	.030	.005	.001	.000	.030	.005	.001	.000
	RMSE	.665	.105	.075	.013	.665	.105	.075	.013	.488	.075	.048	.008	.488	.075	.048	.008
	Coverage	.931	.934	.923	.921	.931	.934	.923	.921	.939	.944	.938	.941	.939	.944	.938	.941
	Time	1409.9	1299.9	1258.7	1272.2	62.6	58.4	57.7	56.7	8675.5	8798.0	8585.9	8809.2	231.8	230.4	229.4	231.8
τ=-0.2	Bias	.033	$-.002$	$-.001$	$-.001$	.033	$-.002$	$-.001$	$-.001$	.013	$-.002$	$-.002$	.001	.013	$-.002$	$-.002$	.001
	RMSE	1.033	.171	.109	.022	1.033	.171	.109	.022	.726	.109	.076	.013	.726	.109	.076	.013
	Coverage	.928	.922	.932	.909	.928	.922	.932	.909	.941	.944	.926	.940	.941	.944	.926	.940
	Time	1514.4	1414.7	1380.0	1506.1	62.6	61.1	60.2	62.3	9806.9	8882.8	8797.3	9949.1	220.1	213.0	212.1	220.7
τ=0.2	Bias	$-.015$	$-.005$	$-.004$	.000	$-.015$	$-.005$	$-.004$	.000	.012	.004	.000	.001	.012	.004	.000	.001
	RMSE	1.038	.147	.112	.021	1.038	.147	.112	.021	.710	.111	.080	.013	.710	.111	.080	.013
	Coverage	.930	.925	.926	.918	.930	.925	.926	.918	.942	.942	.924	.929	.942	.942	.924	.929
	Time	1410.8	1313.5	1302.2	1410.8	60.2	58.9	58.7	60.3	9353.6	8514.3	8419.3	9475.4	218.1	211.3	210.0	218.1
τ=2	Bias	.034	.003	$-.001$	.000	.034	.003	$-.001$	.000	.018	.000	.000	.000	.018	.000	.000	.000
	RMSE	.517	.065	.056	.011	.517	.065	.056	.011	.344	.047	.043	.005	.344	.047	.043	.005
	Coverage	.904	.933	.922	.920	.904	.933	.922	.920	.922	.944	.929	.939	.922	.944	.929	.939
	Time	1421.9	1336.5	1342.3	1344.9	60.1	59.6	59.8	58.1	9333.7	9167.5	8966.0	9179.1	235.0	234.2	232.7	232.8
Results for X_2																	
τ=-2	Bias	$-.009$	$-.002$	.001	.000	$-.009$	$-.002$	.001	.000	.020	.000	$-.002$	.000	.020	.000	$-.002$	.000
	RMSE	.506	.078	.051	.009	.506	.078	.051	.009	.356	.057	.038	.006	.356	.057	.038	.006
	Coverage	.930	.941	.917	.918	.930	.941	.917	.918	.934	.941	.934	.931	.934	.941	.934	.931
	Time	1409.8	1299.9	1258.7	1272.2	62.6	58.4	57.7	56.7	8675.4	8798.0	8585.9	8809.2	231.8	230.3	229.4	231.8
τ=-0.2	Bias	.010	$-.003$	$-.001$	.000	.010	$-.003$	$-.001$	.000	.025	$-.003$	$-.002$	.000	.025	$-.003$	$-.002$	.000
	RMSE	.774	.115	.083	.015	.774	.115	.083	.015	.517	.081	.052	.009	.517	.081	.052	.009
	Coverage	.917	.926	.932	.917	.917	.926	.932	.917	.939	.933	.929	.941	.939	.933	.929	.941
	Time	1514.4	1414.7	1380.0	1506.1	62.6	61.1	60.2	62.3	9807.0	8882.8	8797.3	9949.1	220.1	213.0	212.1	220.7
τ=0.2	Bias	.001	.001	$-.005$	.000	.001	.001	$-.005$	.000	.001	.005	.001	.001	.001	.005	.001	.001
	RMSE	.691	.124	.079	.016	.691	.124	.079	.016	.494	.077	.056	.009	.494	.077	.056	.009
	Coverage	.920	.913	.917	.905	.920	.913	.917	.905	.954	.947	.930	.933	.954	.947	.930	.933
	Time	1410.8	1313.5	1302.2	1410.8	60.2	58.9	58.7	60.3	9353.6	8514.4	8419.3	9475.4	218.1	211.3	210.1	218.2
τ=2	Bias	.019	.001	.000	.000	.019	.001	.000	.000	$-.002$	.003	.001	.000	$-.002$	.003	.001	.000
	RMSE	.378	.045	.052	.007	.378	.045	.052	.007	.211	.038	.028	.005	.211	.038	.028	.005
	Coverage	.909	.928	.929	.910	.909	.928	.929	.910	.938	.942	.941	.929	.938	.942	.941	.929
	Time	1421.9	1336.5	1342.3	1344.9	60.0	59.6	59.8	58.1	9333.7	9167.5	8965.9	9179.1	235.0	234.1	232.6	232.8

Table A.7: The Bayesian results for $\widehat{\beta}_1$ and $\widehat{\beta}_2$

		n=169								n=361							
		expm				mvp				expm				mvp			
		α				α				α				α			
		-2	-0.2	0.2	2	-2	-0.2	0.2	2	-2	-0.2	0.2	2	-2	-0.2	0.2	2
Results for $\widehat{\beta}_1$																	
$\tau=-2$	Bias	−.002	.001	.001	−.016	−.002	.001	.001	−.016	−.007	−.002	−.004	.010	−.007	−.002	−.004	.010
	RMSE	.065	.058	.054	.057	.065	.058	.054	.057	.041	.034	.039	.034	.041	.034	.039	.034
	Coverage	.900	.960	.950	.970	.900	.960	.950	.970	.960	.970	.950	.960	.960	.970	.950	.960
	Time	712.1	657.7	658.3	710.1	26.3	27.3	27.2	25.6	4983.1	4500.9	4502.5	4989.0	36.1	35.1	36.2	34.4
$\tau=-0.2$	Bias	−.007	.000	.002	.002	−.007	.000	.002	.002	−.004	.004	−.002	.002	−.004	.004	−.002	.002
	RMSE	.071	.082	.082	.078	.071	.082	.082	.078	.060	.051	.044	.056	.060	.051	.044	.056
	Coverage	.970	.940	.910	.920	.970	.940	.910	.920	.940	.950	.980	.940	.940	.950	.980	.940
	Time	660.6	604.6	605.1	664.7	26.0	27.1	26.9	25.6	4459.9	4004.1	4004.5	4487.9	36.1	34.3	35.5	34.1
$\tau=0.2$	Bias	−.002	−.005	.011	.005	−.002	−.005	.011	.005	−.004	−.008	.008	−.007	−.004	−.008	.008	−.007
	RMSE	.071	.073	.076	.080	.071	.073	.076	.080	.056	.056	.057	.055	.056	.056	.057	.055
	Coverage	.990	.960	.940	.950	.990	.960	.940	.950	.940	.970	.950	.950	.940	.970	.950	.950
	Time	659.9	608.4	604.9	661.6	26.2	26.3	26.9	26.6	4469.0	4010.0	4004.6	4509.8	36.9	33.2	34.8	34.1
$\tau=2$	Bias	.005	−.001	−.005	.005	.005	−.001	−.005	.005	.000	.001	−.001	.002	.000	.001	−.001	.002
	RMSE	.045	.039	.049	.033	.045	.039	.049	.033	.024	.032	.030	.030	.024	.032	.030	.030
	Coverage	.920	.950	.960	.930	.920	.950	.960	.930	.970	.940	.950	.940	.970	.940	.950	.940
	Time	710.2	656.8	660.0	713.0	26.4	26.8	26.6	25.4	4972.3	4496.2	4508.4	4937.4	36.2	34.1	33.8	34.2
Results for $\widehat{\beta}_2$																	
$\tau=-2$	Bias	−.009	.002	.011	−.014	−.009	.002	.011	−.014	.006	−.001	−.008	−.003	.006	−.001	−.008	−.003
	RMSE	.054	.058	.053	.064	.054	.058	.053	.064	.040	.038	.038	.039	.040	.038	.038	.039
	Coverage	.940	.930	.940	.910	.940	.930	.940	.910	.960	.950	.970	.920	.960	.950	.970	.920
	Time	712.1	657.7	658.3	710.1	26.3	27.3	27.2	25.6	4983.1	4500.9	4502.5	4989.0	36.1	35.1	36.2	34.4
$\tau=-0.2$	Bias	−.010	.006	.005	−.002	−.010	.006	.005	−.002	.010	.006	−.001	.010	.010	.006	−.001	.010
	RMSE	.078	.088	.075	.071	.078	.088	.075	.071	.048	.045	.054	.055	.048	.045	.054	.055
	Coverage	.940	.920	.930	.960	.940	.920	.930	.960	.950	.960	.910	.950	.950	.960	.910	.950
	Time	660.6	604.6	605.1	664.7	26.0	27.1	26.9	25.6	4459.9	4004.1	4004.5	4487.9	36.1	34.3	35.5	34.1
$\tau=0.2$	Bias	.007	.008	.006	−.014	.007	.008	.006	−.014	−.004	.009	.004	−.007	−.004	.009	.004	−.007
	RMSE	.080	.074	.093	.077	.080	.074	.093	.077	.052	.054	.050	.053	.052	.054	.050	.053
	Coverage	.970	.940	.900	.940	.970	.940	.900	.940	.950	.960	.960	.950	.950	.960	.960	.950
	Time	659.9	608.4	604.9	661.6	26.2	26.3	26.9	26.6	4469.0	4010.0	4004.6	4509.8	36.9	33.2	34.8	34.1
$\tau=2$	Bias	−.003	.010	.001	.003	−.003	.010	.001	.003	.004	−.001	−.002	.003	.004	−.001	−.002	.003
	RMSE	.040	.040	.032	.038	.040	.040	.032	.038	.029	.030	.023	.023	.029	.030	.023	.023
	Coverage	.950	.920	.980	.950	.950	.920	.980	.950	.960	.950	.950	.940	.960	.950	.950	.940
	Time	710.2	656.8	660.0	713.0	26.4	26.8	26.6	25.4	4972.3	4496.2	4508.4	4937.4	36.2	34.1	33.8	34.2

Table A.8: The Bayesian results of average direct effects for X_1 and X_2

		n=169								n=361							
		expm				mvp				expm				mvp			
		α				α				α				α			
		−2	−0.2	0.2	2	−2	−0.2	0.2	2	−2	−0.2	0.2	2	−2	−0.2	0.2	2
Results for X_1																	
τ=−2	Bias	−.011	.002	.002	−.028	−.011	.002	.002	−.028	−.008	−.002	−.003	.028	−.008	−.002	−.003	.028
	RMSE	.113	.058	.055	.099	.113	.058	.055	.099	.074	.035	.039	.080	.074	.035	.039	.080
	Coverage	.940	.950	.940	.970	.940	.950	.940	.970	.940	.960	.950	.940	.940	.960	.950	.940
	Time	6623.1	5988.2	5995.2	6603.0	446.7	394.2	395.1	441.1	46853.9	42049.8	41972.0	46883.2	2521.7	2046.1	2060.1	2511.4
τ=−0.2	Bias	−.010	.002	.004	.000	−.010	.002	.004	.000	−.023	.004	−.001	−.006	−.023	.004	−.001	−.006
	RMSE	.143	.082	.084	.148	.143	.082	.084	.148	.121	.052	.044	.107	.121	.052	.044	.107
	Coverage	.970	.950	.910	.950	.970	.950	.910	.950	.920	.950	.980	.980	.920	.950	.980	.980
	Time	6159.4	5540.2	5542.9	6194.0	451.9	392.3	393.6	445.3	42115.0	37546.8	37469.3	42343.4	2513.3	2007.1	2016.9	2508.7
τ=0.2	Bias	−.022	−.003	.013	.001	−.022	−.003	.013	.001	−.017	−.007	.009	−.012	−.017	−.007	.009	−.012
	RMSE	.132	.073	.078	.174	.132	.073	.078	.174	.108	.055	.058	.110	.108	.055	.058	.110
	Coverage	.970	.970	.950	.960	.970	.970	.950	.960	.970	.970	.950	.940	.970	.970	.950	.940
	Time	6153.8	5574.3	5539.1	6161.3	450.2	388.4	393.2	450.4	42169.3	37567.6	37441.1	42523.0	2534.6	1993.2	2012.0	2520.8
τ=2	Bias	.009	−.001	−.004	.005	.009	−.001	−.004	.005	.000	.001	.000	−.001	.000	.001	.000	−.001
	RMSE	.074	.039	.049	.083	.074	.039	.049	.083	.057	.032	.031	.051	.057	.032	.031	.051
	Coverage	.950	.940	.950	.950	.950	.940	.950	.950	.970	.950	.970	.960	.970	.950	.970	.960
	Time	6600.2	5968.9	6003.7	6623.9	446.6	386.7	386.6	440.4	46740.4	41953.4	41968.3	46378.7	2507.7	2005.1	2014.0	2459.6
Results for X_2																	
τ=−2	Bias	−.019	.002	.012	−.025	−.019	.002	.012	−.025	.012	−.001	−.008	.000	.012	−.001	−.008	.000
	RMSE	.092	.059	.053	.106	.092	.059	.053	.106	.068	.039	.038	.065	.068	.039	.038	.065
	Coverage	.950	.930	.930	.930	.950	.930	.930	.930	.960	.950	.970	.940	.960	.950	.970	.940
	Time	6622.9	5988.2	5995.2	6602.9	441.1	388.9	389.8	435.5	46852.5	42047.8	41973.9	46884.1	2520.9	2044.7	2058.7	2509.7
τ=−0.2	Bias	−.016	.006	.006	−.005	−.016	.006	.006	−.005	.009	.006	−.001	.011	.009	.006	−.001	.011
	RMSE	.137	.088	.076	.124	.137	.088	.076	.124	.087	.045	.054	.098	.087	.045	.054	.098
	Coverage	.950	.930	.930	.940	.950	.930	.930	.940	.940	.970	.910	.950	.940	.970	.910	.950
	Time	6158.9	5540.2	5542.7	6193.4	446.3	386.5	388.8	440.0	42113.2	37544.2	37471.8	42345.5	2512.1	2003.5	2012.7	2507.8
τ=0.2	Bias	.001	.009	.007	−.027	.001	.009	.007	−.027	−.012	.010	.004	−.012	−.012	.010	.004	−.012
	RMSE	.136	.075	.094	.143	.136	.075	.094	.143	.092	.054	.050	.093	.092	.054	.050	.093
	Coverage	.950	.940	.900	.960	.950	.940	.900	.960	.960	.970	.960	.940	.960	.970	.960	.940
	Time	6153.1	5574.5	5539.2	6160.9	444.8	382.6	387.8	443.6	42167.6	37565.5	37444.5	42525.4	2533.2	1989.7	2008.4	2519.7
τ=2	Bias	−.006	.010	.001	.004	−.006	.010	.001	.004	.007	−.001	−.002	.004	.007	−.001	−.002	.004
	RMSE	.069	.040	.033	.073	.069	.040	.033	.073	.054	.030	.023	.042	.054	.030	.023	.042
	Coverage	.940	.940	.970	.930	.940	.940	.970	.930	.950	.960	.970	.960	.950	.960	.970	.960
	Time	6599.9	5969.2	6004.0	6623.8	441.5	381.5	381.1	435.7	46739.1	41951.6	41971.0	46380.8	2506.3	2003.2	2011.9	2458.5

Table A.9: The Bayesian results of average indirect effects for X_1 and X_2

		n=169								n=361							
		expm				mvp				expm				mvp			
		α				α				α				α			
		-2	-0.2	0.2	2	-2	-0.2	0.2	2	-2	-0.2	0.2	2	-2	-0.2	0.2	2
Results for X_1																	
τ=-2	Bias	$-.071$	.010	$-.008$	.027	$-.071$	.010	$-.008$	.027	$-.009$	.007	$-.001$	$-.028$	$-.009$	.007	$-.001$	$-.028$
	RMSE	.546	.078	.067	.095	.546	.078	.067	.095	.401	.059	.040	.082	.401	.059	.040	.082
	Coverage	.940	.920	.950	.960	.940	.920	.950	.960	.970	.980	.960	.930	.970	.980	.960	.930
	Time	6968.0	6273.2	6280.8	6948.5	785.4	679.4	681.6	776.6	49273.4	43999.5	43930.4	49319.1	4908.2	3958.2	3985.9	4891.4
τ=-0.2	Bias	$-.018$	.001	$-.002$	.002	$-.018$	.001	$-.002$	.002	$-.161$	$-.004$	.003	.009	$-.161$	$-.004$	.003	.009
	RMSE	.836	.132	.092	.147	.836	.132	.092	.147	.640	.092	.060	.107	.640	.092	.060	.107
	Coverage	.940	.960	.950	.940	.940	.960	.950	.940	.930	.950	.960	1.000	.930	.950	.960	1.000
	Time	6508.1	5827.4	5830.7	6544.0	796.4	675.1	680.5	785.2	44502.3	39476.8	39410.1	44758.0	4892.8	3883.1	3899.8	4887.4
τ=0.2	Bias	$-.157$	.017	$-.004$	.003	$-.157$	.017	$-.004$	.003	$-.112$	.012	$-.002$	.012	$-.112$	.012	$-.002$	.012
	RMSE	.798	.120	.101	.177	.798	.120	.101	.177	.595	.094	.066	.112	.595	.094	.066	.112
	Coverage	.960	.970	.960	.940	.960	.970	.960	.940	.930	.970	.970	.940	.930	.970	.970	.940
	Time	6500.7	5863.8	5827.2	6510.1	793.1	669.1	678.7	790.7	44566.5	39503.8	39383.6	44952.3	4933.2	3857.0	3890.5	4913.1
τ=2	Bias	.030	$-.008$	$-.001$	$-.003$	.030	$-.008$	$-.001$	$-.003$	.004	$-.007$	.000	.002	.004	$-.007$	.000	.002
	RMSE	.352	.068	.060	.086	.352	.068	.060	.086	.355	.046	.036	.050	.355	.046	.036	.050
	Coverage	.950	.980	.940	.940	.950	.980	.940	.940	.930	.980	.950	.960	.930	.980	.950	.960
	Time	6942.3	6252.7	6289.8	6969.0	786.1	665.9	666.2	777.4	49150.4	43891.8	43925.0	48762.6	4880.2	3879.9	3897.7	4789.5
Results for X_2																	
τ=-2	Bias	$-.079$	.006	$-.005$	.023	$-.079$	.006	$-.005$	.023	.052	.004	.001	$-.001$	.052	.004	.001	$-.001$
	RMSE	.370	.043	.033	.098	.370	.043	.033	.098	.284	.031	.021	.061	.284	.031	.021	.061
	Coverage	.960	.950	.950	.930	.960	.950	.950	.930	.960	.990	.950	.940	.960	.990	.950	.940
	Time	6968.5	6275.2	6282.5	6949.3	785.6	679.6	681.8	776.6	49275.7	44000.9	43933.2	49321.9	4911.6	3960.8	3987.6	4893.6
τ=-0.2	Bias	$-.047$	.000	$-.002$	.006	$-.047$	.000	$-.002$	.006	$-.009$	$-.002$	.001	$-.009$	$-.009$	$-.002$	.001	$-.009$
	RMSE	.587	.063	.046	.118	.587	.063	.046	.118	.390	.046	.032	.092	.390	.046	.032	.092
	Coverage	.940	.970	.950	.950	.940	.970	.950	.950	.930	.960	.930	.960	.930	.960	.930	.960
	Time	6508.4	5828.7	5832.1	6544.5	796.7	675.4	680.9	785.2	44504.6	39479.7	39412.0	44761.5	4896.4	3882.8	3900.2	4889.8
τ=0.2	Bias	$-.037$	.012	$-.001$	.027	$-.037$	.012	$-.001$	.027	$-.068$	.009	.000	.011	$-.068$	.009	.000	.011
	RMSE	.572	.067	.051	.137	.572	.067	.051	.137	.404	.050	.033	.089	.404	.050	.033	.089
	Coverage	.910	.970	.940	.960	.910	.970	.940	.960	.960	.960	.950	.930	.960	.960	.950	.930
	Time	6501.3	5865.0	5828.6	6510.6	793.2	669.3	678.6	792.4	44568.6	39506.3	39385.6	44956.1	4935.0	3857.2	3890.8	4914.4
τ=2	Bias	$-.019$	$-.002$	$-.002$	$-.003$	$-.019$	$-.002$	$-.002$	$-.003$	.024	$-.004$	.000	$-.003$	.024	$-.004$	.000	$-.003$
	RMSE	.273	.033	.033	.071	.273	.033	.033	.071	.243	.024	.018	.041	.243	.024	.018	.041
	Coverage	.910	.960	.950	.930	.910	.960	.950	.930	.940	.930	.970	.960	.940	.930	.970	.960
	Time	6942.9	6254.7	6291.6	6970.0	786.4	666.2	666.2	777.5	49152.6	43894.2	43927.7	48764.1	4882.3	3881.5	3899.1	4792.2

Table A.10: The Bayesian results of average total effects for X_1 and X_2

		n=169								n=361							
		expm				mvp				expm				mvp			
		α				α				α				α			
		-2	-0.2	0.2	2	-2	-0.2	0.2	2	-2	-0.2	0.2	2	-2	-0.2	0.2	2
Results for X_1																	
τ=-2	Bias	$-.082$	.011	$-.006$	$-.001$	$-.082$	.011	$-.006$	$-.001$	$-.018$	.005	$-.004$	.000	$-.018$	.005	$-.004$	.000
	RMSE	.647	.103	.073	.012	.647	.103	.073	.012	.467	.081	.047	.007	.467	.081	.047	.007
	Coverage	.940	.920	.970	.930	.940	.920	.970	.930	.970	.940	.980	.950	.970	.940	.980	.950
	Time	6622.8	5988.8	5996.0	6602.9	440.6	388.3	389.4	435.0	46847.3	42042.2	41970.5	46880.4	2516.1	2040.8	2054.6	2505.6
τ=-0.2	Bias	$-.029$	.003	.002	.002	$-.029$	.003	.002	.002	$-.185$	.000	.002	.003	$-.185$	.000	.002	.003
	RMSE	.965	.168	.096	.019	.965	.168	.096	.019	.753	.117	.071	.013	.753	.117	.071	.013
	Coverage	.960	.950	.980	.940	.960	.950	.980	.940	.930	.930	.980	.960	.930	.930	.980	.960
	Time	6158.9	5540.6	5543.4	6193.5	445.9	386.0	388.3	439.3	42109.2	37538.8	37467.6	42341.2	2507.6	2000.0	2009.6	2503.9
τ=0.2	Bias	$-.180$	.014	.010	.004	$-.180$	.014	.010	.004	$-.129$	.005	.008	.000	$-.129$	.005	.008	.000
	RMSE	.914	.147	.107	.020	.914	.147	.107	.020	.694	.111	.070	.013	.694	.111	.070	.013
	Coverage	.960	.980	.960	.950	.960	.980	.960	.950	.930	.970	.970	.960	.930	.970	.970	.960
	Time	6153.1	5575.0	5539.8	6160.9	444.3	382.1	387.2	443.1	42164.0	37560.7	37441.2	42521.2	2529.1	1986.4	2004.5	2515.6
τ=2	Bias	.038	$-.009$	$-.005$	.002	.038	$-.009$	$-.005$	.002	.004	$-.006$	$-.001$	.001	.004	$-.006$	$-.001$	.001
	RMSE	.416	.082	.074	.009	.416	.082	.074	.009	.408	.058	.039	.007	.408	.058	.039	.007
	Coverage	.940	.920	.950	.920	.940	.920	.950	.920	.940	.940	.960	.960	.940	.940	.960	.960
	Time	6599.8	5969.7	6004.6	6623.9	441.0	380.7	380.5	435.2	46734.2	41946.5	41968.2	46376.3	2502.4	1999.5	2008.0	2454.7
Results for X_2																	
τ=-2	Bias	$-.097$	.008	.006	$-.001$	$-.097$	.008	.006	$-.001$	.065	.003	$-.007$	$-.001$	.065	.003	$-.007$	$-.001$
	RMSE	.457	.087	.057	.010	.457	.087	.057	.010	.348	.059	.036	.007	.348	.059	.036	.007
	Coverage	.960	.920	.930	.890	.960	.920	.930	.890	.960	.940	.970	.960	.960	.940	.970	.960
	Time	6622.1	5989.0	5995.9	6602.3	439.8	387.5	388.5	434.2	46845.4	42038.7	41967.7	46877.6	2513.3	2037.6	2051.9	2502.8
τ=-0.2	Bias	$-.063$	.006	.004	.001	$-.063$	.006	.004	.001	.000	.004	.000	.003	.000	.004	.000	.003
	RMSE	.717	.115	.075	.012	.717	.115	.075	.012	.472	.070	.050	.009	.472	.070	.050	.009
	Coverage	.940	.940	.950	.950	.940	.940	.950	.950	.950	.950	.970	.950	.950	.950	.970	.950
	Time	6158.1	5540.3	5543.2	6192.8	444.9	385.2	387.6	438.5	42107.0	37535.4	37466.3	42339.5	2505.1	1999.1	2009.4	2500.6
τ=0.2	Bias	$-.036$	.021	.006	.000	$-.036$	.021	.006	.000	$-.080$	.019	.004	.000	$-.080$	.019	.004	.000
	RMSE	.698	.120	.092	.013	.698	.120	.092	.013	.492	.082	.051	.009	.492	.082	.051	.009
	Coverage	.920	.950	.940	.940	.920	.950	.940	.940	.980	.960	.970	.950	.980	.960	.970	.950
	Time	6152.3	5574.7	5539.6	6160.3	443.4	381.0	386.3	442.1	42161.3	37558.1	37439.3	42518.5	2526.6	1985.5	2003.5	2512.8
τ=2	Bias	$-.025$	.008	.000	.001	$-.025$	.008	.000	.001	.031	$-.005$	$-.001$	.001	.031	$-.005$	$-.001$	.001
	RMSE	.338	.053	.032	.006	.338	.053	.032	.006	.294	.043	.026	.004	.294	.043	.026	.004
	Coverage	.920	.930	.950	.970	.920	.930	.950	.970	.930	.940	.930	.950	.930	.940	.930	.950
	Time	6599.4	5970.1	6004.9	6623.3	440.1	380.1	379.6	434.2	46731.9	41943.0	41965.9	46373.9	2498.8	1997.6	2005.7	2451.6