HAL open science

Large sample properties of the matrix exponential spatial specification with an application to FDI

Nicolas Debarsy, Fei Jin, Lung-Fei Lee

To cite this version:

Nicolas Debarsy, Fei Jin, Lung-Fei Lee. Large sample properties of the matrix exponential spatial specification with an application to FDI. 2014. hal-01069198

HAL Id: hal-01069198

https://hal.science/hal-01069198v1

Preprint submitted on 28 Sep 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Large sample properties of the matrix exponential spatial specification with an application to FDI

First version: September 2013
Current version: September 2014

Nicolas Debarsy[a,1], Fei Jin[b], Lung-fei Lee[c]

[a] *Univ. Orléans, CNRS, LEO, UMR 7322, F45067, Orléans, France*
[b] *School of Economics, Shanghai University of Finance and Economics, Shanghai 200433, China*
[c] *Department of Economics, The Ohio State University, Columbus, OH 43210, USA*

Abstract

This paper studies large sample properties of the matrix exponential spatial specification (MESS). We find that the quasi-maximum likelihood estimator (QMLE) for the MESS is consistent under heteroskedasticity, a property not shared by the QMLE of the SAR model. For the general model that has MESS in both the dependent variable and disturbances, labeled MESS(1,1), the QMLE can be consistent under unknown heteroskedasticity when the spatial weights matrices in the two MESS processes are commutative. We also consider the generalized method of moments estimator (GMME). In the homoskedastic case, we derive a best GMME that is as efficient as the maximum likelihood estimator under normality and can be asymptotically more efficient than the QMLE under non-normality. In the heteroskedastic case, an optimal GMME can be more efficient than the QMLE asymptotically. The QML approach for the MESS model has the computational advantage over that of a SAR model. The computational simplicity carries over to MESS models with any finite order of spatial matrices. No parameter range needs to be imposed in order for the model to be stable. Results of Monte Carlo experiments for finite sample properties of the estimators are reported. Finally, the MESS(1,1) is applied to Belgium's outward FDI data and we observe that the dominant motivation of Belgium's outward FDI lies in finding cheaper factor inputs.

Keywords: Spatial autocorrelation, MESS, QML, GMM, Heteroskedasticity, Delta method, FDI
JEL Classification: C12, C13, C21, F14, F21

Email addresses: `nicolas.debarsy@univ-orleans.fr` (Nicolas Debarsy), `jin.fei@shufe.edu.cn` (Fei Jin), `lee.1777@osu.edu` (Lung-fei Lee)

[1] Research associate of the CNRS.

1. Introduction

The Matrix Exponential Spatial Specification (MESS) has been initially proposed by LeSage and Pace (2007) as a substitute to the well-known spatial autoregressive (SAR) specification. The difference between the two rests on the type of decay which characterizes the influence of space. The MESS uses an exponential decay while the SAR specification is based on a geometrical decay. The motivation of these authors to use the MESS is its computational simplicity. Indeed, in contrast to the SAR, the quasi-maximum likelihood (QML) function of the MESS does not involve any Jacobian of the transformation and thus reduces to a nonlinear regression estimation. This is so even for its extension to models with a finite number of spatial weights matrices. A second advantage of the MESS is the absence of constraints on the parameter space of the coefficient that captures interactions between observations since the reduced form of the MESS always exists (see Chiu et al., 1996). Furthermore, no positivity constraint on the Jacobian of the transformation needs to be imposed as it does not appear in the quasi log-likelihood function. In Section 2, we nevertheless show that MESS and SAR models cannot be seen as perfect substitutes since neither a one-to-one correspondence between the parameters capturing interactions nor between impacts (except in some specific cases) can be derived. Furthermore, a MESS model is always a stable spatial process, but a SAR model with strong spatial interaction might be unstable.[2]

A third advantage of the MESS, proved in this paper, is that the quasi-maximum likelihood estimator (QMLE) is consistent even in the presence of unknown heteroskedasticity, a feature not shared by the SAR model (see Lin and Lee, 2010, p. 36). These two authors have however shown in this SAR context that a Generalized Method of Moments Estimator (GMME) with properly modified quadratic moment conditions could still be consistent in presence of unknown heteroskedasticity.[3] Using quadratic moment conditions similar to those in Lin and Lee (2010), we derive an optimal GMME consistent in presence of unknown heteroskedasticity and also generally more efficient with respect to the QMLE (with either normal or non-normal disturbances). The relative efficiency of the optimal GMME results from the optimal weighting of the GMM estimation method which uses the same moments that the QMLE integrates. In the homoskedastic case, we derive a best (optimal) GMME that is as efficient as the MLE under normality and can be more efficient than the QMLE under non-normality.[4] The best GMME takes a much simpler form than that for

[2] From this view, we may argue that the MESS would be useful only when observed outcomes do not show unstable phenomena.

[3] Kelejian and Prucha (2010) also develop a GMME robust to the presence of heteroskedasticity but their main focus is on spatial autocorrelation in the error terms.

[4] Lee (2007) derives the best optimal GMME for the SAR model with normal i.i.d. disturbances, which is as efficient as the QMLE. Liu et al. (2010) consider the best optimal GMME for the SAR model with SAR disturbances that can be more efficient than the QMLE under non-normality, which is extended to high order SAR models in Lee and Liu (2010).

the SAR model and the optimal orthogonal conditions do not involve any estimated parameters[5].

Even though LeSage and Pace (2007) present the maximum likelihood and Bayesian estimators of the MESS, no asymptotic theory has been derived for this specification. In this paper, we focus our attention on the general model where a MESS is present in both the dependent variable and in the error terms (MESS(1,1) for short), and develop large sample properties for QML and GMM methods under both homoskedastic and heteroskedastic cases.[6] In the homoskedastic case, the best GMME for models with normal disturbances or commutative spatial weights matrices in the MESS(1,1) is as efficient as the QMLE but generally more efficient than the QMLE for other ones. In the (unknown) heteroskedastic case, the QMLE for the MESS(1,1) can be consistent only when the spatial weights matrices for the MESS in the dependent variable and in disturbances are commutative, but it is less efficient than an optimal GMME. If different variances in the heteroskedastic case could be estimated consistently, a best GMME could also be implemented.[7] We also perform Monte Carlo experiments to assess the small sample performance of our proposed estimators.

Analysis of significance of determinants' causal effects on the dependent variable is of interest for economists. In this paper, we derive a lemma allowing to perform inference on the elements of the matrix of impacts implied by the reduced form of the MESS(1,1). The lemma is based on an adapted version of the Delta method and can be used to test the significance of (functions of) impacts as long as the number of constraints is not dependent on the sample size. This lemma is valuable for applied economists since until now, with the exception of LeSage and Pace (2009) who provide inference for scalar summaries of these impacts in the SAR model either by simulating the distributions or estimating them via Bayesian methods,[8] there does not exist any classical statistical test to assess the significance of (functions of) individual impacts.

The developed estimators are finally applied to a modified gravity equation aimed at explaining Belgium's outward FDI. Blonigen et al. (2007) propose four different classifications of FDI which can be distinguished based on the sign of spatial autocorrelation and market-potential of host countries. In addition to obtaining significant and expected signs for the traditional variables included in the gravity model when spatial autocorrelation is accounted for, namely GDP, population and bilateral distance, we find a significant negative spatial autocorrelation and a positive but non-significant market potential effect for hosts countries. Thus vertical FDI is the dominant type of outward FDI for Belgium. We further compare MESS(1,1) and

[5] See Lee (2003, 2007) for further details.

[6] In a supplementary file, we consider the QML estimation of a high order MESS, namely MESS(p,q), with p and q being the orders of the MESS in the dependent variable and in the errors respectively. While the parameter spaces for high order SAR models can be hard to find (Lee and Liu, 2010; Elhorst et al., 2012), high order MESS models have the advantage that the parameter spaces are not restricted.

[7] For the SAR model under unknown heteroskedasticity, Lin and Lee (2010) have not discussed the possible best GMME.

[8] See Elhorst (2010) for a step-by-step explanation on how to simulate the distributions.

SARAR results and show that their economic conclusions in terms of impacts are very similar. However, the MESS, for several reasons, namely computational, technical and statistical, can be more appealing. Finally, statistical significance on impacts is analyzed through the application of the derived lemma for inference.

The rest of the paper is organized as follows. Section 2 compares MESS and SAR models in a more formal way. Section 3 considers the large sample properties the QML and GMM estimators under both homoskedasticity and unknown heteroskedasticity. It also derives a lemma to perform inference on the elements of the matrix of impacts of explanatory variables obtained from the reduced form of the MESS(1,1). Section 4 presents Monte Carlo experiments while Section 5 presents the application of our estimators and applies the lemma for inference on the determinants of Belgium's outward FDI. Section 6 concludes. Some lemmas and proofs are collected in Appendix A.[9]

2. Comparison of MESS and SAR Specifications

The MESS in LeSage and Pace (2007) is

$$e^{\alpha W_n} y_n = X_n \beta + \epsilon_n, \quad \epsilon_n = (\epsilon_{n1}, \ldots, \epsilon_{nn})', \tag{1}$$

where n is the sample size, y_n is an n-dimensional vector of observations on the dependent variable, X_n is an $n \times k$ matrix of exogenous variables with corresponding coefficient vector β, W_n is an $n \times n$ spatial weights matrix modeling interactions among observations (with zero diagonal elements), ϵ_{ni}'s are independent with mean zero, and α is the parameter measuring the intensity of interactions between observations. For any $n \times n$ square matrix A_n, let A_n^0 be the $n \times n$ identity matrix I_n. The matrix exponential $e^{\alpha A_n}$, defined as $e^{\alpha A_n} = \sum_{i=0}^{\infty} \frac{\alpha^i A_n^i}{i!}$, is always invertible, with the inverse being $e^{-\alpha A_n}$ (Chiu et al., 1996). As a result, the variance-covariance (VC) matrix of y_n which equals to $e^{-\alpha_0 W_n} \mathrm{E}(\epsilon_n \epsilon_n') e^{-\alpha_0 W_n'}$ with α_0 being the true value of α, is always positive definite. No restriction on the parameter space of α needs to be imposed.

In this paper, we consider a general model that has MESS in both the dependent variable and the disturbances that we label MESS(1,1) (which should be viewed as an analog of the SAR model with SAR disturbances, i.e., SARAR model):[10]

$$e^{\alpha W_n} y_n = X_n \beta + u_n, \quad e^{\tau M_n} u_n = \epsilon_n, \quad \epsilon_n = (\epsilon_{n1}, \ldots, \epsilon_{nn})', \tag{2}$$

[9] Except the proof of Proposition 8 which is presented in Appendix A, proofs of remaining propositions are similar to those in Lee (2004) and Lee (2007). Those proofs are provided in a supplementary file, which is available upon request.

[10] As pointed out by an anonymous referee, on the r.h.s. of the main equation of (2), to reflect local spatial dependence as in a spatial Durbin model, we may include an additional term $W_n X_n$, or $W_n X_{1n}$ if W_n is row-normalized and X_n contains an intercept term so that $X_n = [l_n, X_{1n}]$. However, as $W_n X_n$ or $W_n X_{1n}$ has the same properties as X_n, the asymptotic analyses below will be similar. Thus, the additional term is not included for simplicity.

where W_n and M_n are $n \times n$ spatial weights matrices. The M_n may or may not be different from W_n. For purposes of comparison and later reference, we put down the SARAR model with the same W_n, M_n, X_n, y_n and ϵ_n:

$$(I_n - \lambda W_n)y_n = X_n\beta + u_n, \quad (I_n - \rho M_n)u_n = \epsilon_n. \tag{3}$$

The parameter spaces of λ and ρ should be restricted so that the VC matrix of y_n, namely $(I_n - \lambda W_n)^{-1}(I_n - \rho M_n)^{-1}\,\mathrm{E}\,(\epsilon_n\epsilon_n')\,(I_n - \rho M_n')^{-1}(I_n - \lambda W_n')^{-1}$ exists. For the QMLE of the SARAR model with a normalized W_n matrix, the parameter space for λ is typically considered to be $(-1, 1)$.[11]

The quasi log likelihood function of the MESS(1,1) presented in (2), as if the ϵ_{ni}'s were i.i.d. normal, is

$$L_n(\theta) = -\frac{n}{2}\ln(2\pi\sigma^2) + \ln|e^{\alpha W_n}| + \ln|e^{\tau M_n}| - \frac{1}{2\sigma^2}(e^{\alpha W_n}y_n - X_n\beta)'e^{\tau M_n'}e^{\tau M_n}(e^{\alpha W_n}y_n - X_n\beta),$$

where $\theta = (\gamma', \sigma^2)'$ with $\gamma = (\alpha, \tau, \beta')'$. Let θ_0 be the true parameter vector. Since $|e^{\alpha W_n}| = e^{\alpha\,\mathrm{tr}(W_n)}$ and $|e^{\tau M_n}| = e^{\tau\,\mathrm{tr}(M_n)}$, as long as W_n and M_n have zero diagonals, the Jacobian of the transformation disappears and the quasi log likelihood function is simplified to

$$L_n(\theta) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(e^{\alpha W_n}y_n - X_n\beta)'e^{\tau M_n'}e^{\tau M_n}(e^{\alpha W_n}y_n - X_n\beta). \tag{4}$$

By contrast, the quasi log likelihood function of the SARAR model shown in (3) involves the log determinant of the Jacobian $\ln|(I_n - \lambda W_n)(I_n - \rho M_n)| = \ln|I_n - \lambda W_n| + \ln|I_n - \rho M_n|$, which may make the QMLE computationally intensive for large sample sizes.

Another difference between these two specifications is that one does not need to normalize the interaction matrices in the MESS. In the SARAR model, the purpose of normalizing the interaction matrices is to standardize the parameter spaces for λ and ρ so that they correspond to $(-1, 1)$, which facilitates the interpretation of these parameters. However, in the MESS, since no parameter constraint is involved, the normalization of the interaction matrices may not play a special role.

LeSage and Pace (2007) present the MESS as a computationally simpler substitute for the SAR model.[12] Using a row-normalized interaction matrix W_n, they propose the approximated relation $\lambda = 1 - e^{\alpha}$. They argue that this approximation is derived by equating the length of $||e^{\alpha W_n}||_\infty$ and $||I_n - \lambda W_n||_\infty$, with $||.||_\infty$ being the maximum row sum matrix norm. However, this approximation is not always right since the matrix

[11] See Kelejian and Prucha (2010) for a detailed discussion about the parameter space for λ. For high order SARAR models, finding the parameter spaces can be hard. Elhorst et al. (2012) have outlined a procedure for finding the stationary region, but the parameter spaces can be complicated even for a SAR model with two spatial weights matrices for the dependent variable. By contrast, parameters in high order MESS models, labeled MESS(p,q), where p and q are the orders of the MESS in the dependent variable and disturbances respectively, do not need to be restricted and the effort to find the parameter spaces is saved. The supplementary file considers the QML estimation of these high order MESS models.

[12] Han and Lee (2012) consider the J-test procedure to choose between MESS and SAR models.

norm should be taken over the absolute value of matrix elements. By contrast, if one turns to the impact analysis, an equivalence between the two specifications can be traced back at least in some specific cases.

Before presenting this correspondence, it is important to discuss the features of impact analysis in spatial autoregressive (SAR or MESS) regressions. Impact analysis, which is one of the main focuses for economists, is based on the reduced form of the estimated econometric specification. For the MESS(1,1) case, the reduced form is $y_n = e^{-\alpha W_n}(X_n\beta + e^{-\tau M_n}\epsilon_n)$. One then computes the matrix of impact for each regressor X_{nk}, $k = 1, \cdots, k$, by calculating the partial derivative of y_n with respect to the concerned regressor. For a continuous regressor X_{nk}, this matrix is

$$\Xi^{y_n}_{X_{nk}} = \frac{\partial E(y_n|X_n)}{\partial X'_{nk}} = \beta_k e^{-\alpha W_n}. \tag{5}$$

The diagonal elements of this matrix contain the direct effects including own-spillover effects, which are inherently heterogeneous in presence of spatial autocorrelation due to differentiated friction terms in the interaction matrix. This is what Debarsy and Ertur (2010) call interactive heterogeneity. Off-diagonal elements of this matrix represent indirect effects, meaning the impact of a change in explanatory variable for individual j on the dependent variable for individual i. These direct and indirect effects are, respectively, $\frac{\partial E(y_n|X_n)i}{\partial X_{nk,i}} \equiv \left(\Xi^{y_n}_{X_{nk}}\right)_{ii}$ and $\frac{\partial E(y_n|X_n)i}{\partial X_{nk,j}} \equiv \left(\Xi^{y_n}_{X_{nk}}\right)_{ij}$. For the SARAR model, its associated reduced form is $y_n = (I_n - \lambda W_n)^{-1}[X_n\beta + (I_n - \rho M_n)^{-1}\epsilon_n]$, and the implied impact matrix for regressor X_{nk} is

$$\Xi^{y_n}_{X_{nk}} = \frac{\partial E(y_n|X_n)}{\partial X'_{nk}} = \beta_k (I_n - \lambda W_n)^{-1}. \tag{6}$$

To summarize the information conveyed by these matrices of impacts, LeSage and Pace (2009) propose extracting several scalar measures, as the average direct effect (mean of the diagonal elements), average total effect (average of the row or column sums) and average indirect effect (average of the column or row sums excluding the diagonal element).

Consider a row-normalized interaction matrix W_n in the MESS(1,1) model. Suppose that a shock of the same magnitude Δx is applied on the kth explanatory variable X_{nk} to all spatial units. The new explanatory variable is now $X_{nk} + l_n\Delta x$, with l_n being the n-dimensional vector of ones. For the MESS(1,1), from its reduced form, one calculates a total impact of $\Delta y_n = e^{-\alpha W_n} l_n \Delta x \beta_k$. The average total effect is thus equal to $\frac{1}{n} l'_n \Delta y_n = e^{-\alpha}\Delta x \beta_k$.[13] Correspondingly, the average total impact of X_{nk} in the SARAR model is $\frac{1}{n} l'_n \Delta y_n = \frac{1}{1-\lambda}\Delta x \beta_k$. Equating the two gives the relation $\alpha = \ln(1-\lambda)$ or $\lambda = 1 - e^{\alpha}$. Thus, there is a negative relation between λ and α. $\lambda = 0$ if and only if $\alpha = 0$. When $0 < \lambda < 1$, α will take on negative values and vice-versa. When the normalization used for W_n differs from the row-normalization,

[13] As W_n is row-normalized, $W_n^k l_n = W_n l_n = l_n$, $k \in \mathbb{N}$.

such a relation does not exist.

Even though a relation between λ and α can be found for a row-normalized W_n, we nevertheless cannot consider these two models as substitutes of each other. The underlying reason lies in the comparison of parameter spaces. As mentioned above, for the SARAR model with normalized W_n, λ is usually restricted to the range $(-1, 1)$. However, in the MESS(1,1), $\alpha \in (-\infty, \infty)$. So, while $\lambda < -1$ is not allowed for a SARAR model, α can be greater than ln(2), meaning that parameter spaces of α and λ do not correspond. So, for high negative spatial autocorrelation, we could observe substantial difference between these two models.[14] Furthermore, in a SAR model, if $\lambda > 1$, it would be an unstable model, while unstability does not occur for the MESS with any finite value of α.

3. Estimations of the MESS(1,1) Model

We consider the QML estimation as well as the GMM estimation of the MESS(1,1) in this section. From (4), it is apparent that the QMLE of γ is the minimizer of the function

$$Q_n(\gamma) = (e^{\alpha W_n} y_n - X_n\beta)' e^{\tau M_n'} e^{\tau M_n} (e^{\alpha W_n} y_n - X_n\beta). \tag{7}$$

The derivatives of $Q_n(\gamma)$ with respect to α, τ and β at γ_0 are, respectively,

$$\frac{\partial Q_n(\gamma_0)}{\partial \alpha} = 2(X_n\beta_0 + e^{-\tau_0 M_n}\epsilon_n)' W_n' e^{\tau_0 M_n'}\epsilon_n, \quad \frac{\partial Q_n(\gamma_0)}{\partial \tau} = 2\epsilon_n' M_n \epsilon_n, \quad \frac{\partial Q_n(\gamma_0)}{\partial \beta} = -2X_n' e^{\tau_0 M_n'}\epsilon_n. \tag{8}$$

When ϵ_{ni}'s are i.i.d. with mean zero and variance σ_0^2, as $\mathrm{E}(\epsilon_n' M_n \epsilon_n) = \mathrm{tr}[M_n \mathrm{E}(\epsilon_n\epsilon_n')] = \sigma_0^2 \mathrm{tr}(M_n) = 0$ and $\mathrm{E}(\epsilon_n' e^{-\tau_0 M_n'} W_n' e^{\tau_0 M_n'}\epsilon_n) = \sigma_0^2 \mathrm{tr}(W_n' e^{\tau_0 M_n'} e^{-\tau_0 M_n'}) = \sigma_0^2 \mathrm{tr}(W_n') = 0$, the expected value of $\frac{\partial Q_n(\gamma_0)}{\partial \gamma}$ is zero, which verifies that the minimizer of $\mathrm{E}\, Q_n(\gamma)$ can be γ_0. When ϵ_{ni}'s are independent with mean zero but different variances σ_{ni}^2's, $\mathrm{E}(\epsilon_n' M_n \epsilon_n) = \mathrm{tr}(M_n \Sigma_n) = 0$ since the diagonal elements of M_n are all zero, and $\Sigma_n = \mathrm{Diag}(\sigma_{n1}^2, \ldots, \sigma_{nn}^2)$ is a diagonal matrix containing the different variances as diagonal elements. In addition, $\mathrm{E}(\epsilon_n' e^{-\tau_0 M_n'} W_n' e^{\tau_0 M_n'}\epsilon_n) = \mathrm{tr}(e^{-\tau_0 M_n'} W_n' e^{\tau_0 M_n'} \Sigma_n)$, which may not be zero in general. But if $W_n M_n = M_n W_n$, then $W_n' e^{\tau_0 M_n'} = e^{\tau_0 M_n'} W_n'$ and $\mathrm{E}(\epsilon_n' e^{-\tau_0 M_n'} W_n' e^{\tau_0 M_n'}\epsilon_n) = \mathrm{tr}(W_n'\Sigma_n) = 0$. Therefore, when the matrix W_n in the spatial lag process can be commutative with the matrix M_n in the spatial error process, the QMLE for γ, derived from the minimization of $Q_n(\gamma)$, can be consistent even under unknown heteroskedasticity. This includes the special cases that there is no MESS process in the disturbances or that $M_n = W_n$. This robustness of the QMLE for the MESS(1,0) and MESS(1,1) to unknown heteroskedasticity is a nice feature not shared by the QMLE for the SARAR model.

[14] For a non-negative and row-normalized symmetric interaction matrix W_n, the parameter space for λ may be taken as the interval $(\frac{1}{\mu_{\min,n}}, 1)$ with $\mu_{\min,n}$ being the minimal real eigenvalue of W_n. However, it does not change the conclusions regarding the difference between parameter spaces for λ and α.

The function $Q_n(\gamma)$ may be written as $Q_n(\gamma) = (y_n - e^{-\alpha W_n} X_n\beta)'(e^{-\alpha W_n} e^{-\tau M_n} e^{-\tau M_n'} e^{-\alpha W_n'})^{-1}(y_n - e^{-\alpha W_n} X_n\beta)$. Using the reduced form of the MESS(1,1), namely $y_n = e^{-\alpha_0 W_n}(X_n\beta_0 + e^{-\tau_0 M_n}\epsilon_n)$, and assuming that $\mathrm{E}(\epsilon_n\epsilon_n') = \sigma_0^2 I_n$, the VC matrix of y_n is $\sigma_0^2 e^{-\alpha_0 W_n} e^{-\tau_0 M_n} e^{-\tau_0 M_n'} e^{-\alpha_0 W_n'}$ and the QMLE can be seen as a continuously updating version of the generalized nonlinear least squares (GNLS). The similarity between the QML and GNLS is due to the special structure of the matrix exponential specification. By contrast, there is no such a similarity for the SARAR model (3).[15]

In addition to the QML estimation, we may also consider the GMM estimation of the MESS(1,1) using both linear and quadratic moments, as for the SARAR model. The linear moments would be of the form $F_n'\epsilon_n(\gamma) = 0$, where $\epsilon_n(\gamma) = e^{\tau M_n}(e^{\alpha W_n} y_n - X_n\beta)$ and F_n is an $n \times k_f$ matrix of instruments constructed as functions of W_n and X_n, as in the two-stage least squares (2SLS) approach. The quadratic moments have the form $\epsilon_n'(\gamma)P_{ni}\epsilon_n(\gamma) = 0$; $i = 1, \cdots, k_p$, where P_{ni} has trace zero when ϵ_{ni}'s are i.i.d., implying that $\mathrm{E}[\epsilon_n'(\gamma_0)P_{ni}\epsilon_n(\gamma_0)] = \sigma_0^2 \operatorname{tr}(P_{ni}) = 0$. On the other hand, if the diagonal elements of P_{ni} are all zero when ϵ_{ni}'s are independently distributed with possibly different variances, we get $\mathrm{E}[\epsilon_n'(\gamma_0)P_{ni}\epsilon_n(\gamma_0)] = \operatorname{tr}(P_{ni}\Sigma_n) = 0$.

The basic regularity conditions for estimation are assumed below. The specific sets of hypotheses required for both methods will be given subsequently.

Assumption 1. *Matrices $\{W_n\}$ and $\{M_n\}$ are bounded in both row and column sum norms. The diagonal elements of W_n and M_n are zero.*

Assumption 2. *Elements of X_n are uniformly bounded constants, X_n has full column rank, and besides, $\lim_{n\to\infty} X_n'X_n/n$ exists and is nonsingular.*

Assumptions 1 and 2 follow from the literature, see, e.g., Kelejian and Prucha (1998) and Lee (2004).

3.1. QMLE

For the QMLE from (4), we may just investigate the minimizers of the functions $\{Q_n(\gamma)\}$. For a fixed $\phi = (\alpha, \tau)'$, minimizing $Q_n(\gamma)$ yields $\hat\beta_n(\phi) = (X_n' e^{\tau M_n'} e^{\tau M_n} X_n)^{-1} X_n' e^{\tau M_n'} e^{\tau M_n} e^{\alpha W_n} y_n$. Substituting $\hat\beta_n(\phi)$ into $Q_n(\gamma)$, we obtain a function of only ϕ:

$$Q_n(\phi) = y_n' e^{\alpha W_n'} e^{\tau M_n'} H_n(\tau) e^{\tau M_n} e^{\alpha W_n} y_n, \tag{9}$$

[15] A function for the SARAR model with a structure similar to $Q_n(\gamma)$ is $Q_n^*(\psi) = [(I_n - \lambda W_n)y_n - X_n\beta]'(I_n - \rho M_n')(I_n - \rho M_n)[(I_n - \lambda W_n)y_n - X_n\beta]$ with $\psi = (\lambda, \rho, \beta')'$. At ψ_0, $\mathrm{E}(\frac{\partial Q_n^*(\psi_0)}{\partial\lambda}) = -2\sigma_0^2 \operatorname{tr}[W_n(I_n - \lambda_0 W_n)^{-1}]$ and $\mathrm{E}(\frac{\partial Q_n^*(\psi_0)}{\partial\rho}) = -2\sigma_0^2 \operatorname{tr}[M_n(I_n - \rho_0 M_n)^{-1}]$, where $\operatorname{tr}[W_n(I_n - \lambda_0 W_n)^{-1}] \neq 0$ and $\operatorname{tr}[M_n(I_n - \rho_0 M_n)^{-1}] \neq 0$ in general. Thus the minimizer of $Q_n^*(\psi)$ is not expected to be a consistent estimator of ψ in the SARAR model (3).

where $H_n(\tau) = I_n - e^{\tau M_n} X_n (X_n' e^{\tau M_n'} e^{\tau M_n} X_n)^{-1} X_n' e^{\tau M_n'}$ is a projection matrix. The function $Q_n(\phi)$ can be used for the analysis of the consistency of the QMLE. Although we may not need to restrict the parameter space of ϕ in practice, ϕ should be bounded in analysis so that $e^{\alpha W_n}$ and $e^{\tau M_n}$ would be bounded in both row and column sum norms, since $||e^{\alpha W_n}|| = ||\sum_{i=0}^{\infty} \frac{\alpha^i W_n^i}{i!}|| \leq \sum_{i=0}^{\infty} \frac{|\alpha|^i ||W_n||^i}{i!} = e^{|\alpha| \cdot ||W_n||}$, which is bounded if α is bounded, and so is $||e^{\tau M_n}||$ if τ is bounded, where $||\cdot||$ denotes either the row or column sum matrix norm.

Assumption 3. *There exists a constant $\delta > 0$ such that $|\alpha| \leq \delta$, $|\tau| \leq \delta$ and the true ϕ_0 is in the interior of the parameter space $\Phi = [-\delta, \delta] \times [-\delta, \delta]$.*

For consistency of the QMLE, we need to show that the difference between $Q_n(\phi)/n$ and some non-stochastic function $\bar{Q}_n(\phi)/n$ converges to zero uniformly over the parameter space Φ.[16] The $\bar{Q}_n(\phi)$ will have different forms in the homoskedastic and heteroskedastic cases. By Assumptions 2 and 3, $\frac{1}{n} X_n' e^{\tau M_n'} e^{\tau M_n} X_n$ is bounded. The $Q_n(\phi)$ is a well-defined function for large enough n if the limit of $\frac{1}{n} X_n' e^{\tau M_n'} e^{\tau M_n} X_n$ exists and is nonsingular. In addition, we require that the sequence of the smallest eigenvalues of $e^{\tau M_n'} e^{\tau M_n}$ be bounded away from zero uniformly in τ, so that $H_n(\tau)$ is bounded in both row and column sum norms uniformly in τ. As $e^{\tau M_n'} e^{\tau M_n}$ is positive definite, its smallest eigenvalue is positive. The condition further limits all the eigenvalues to be strictly positive uniformly over the parameter space for all n.

Assumption 4. *The limit $\lim_{n\to\infty} \frac{1}{n} X_n' e^{\tau M_n'} e^{\tau M_n} X_n$ exists and is nonsingular for any $\tau \in [-\delta, \delta]$, and the sequence of the smallest eigenvalues of $e^{\tau M_n'} e^{\tau M_n}$ is bounded away from zero uniformly in $\tau \in [-\delta, \delta]$.*

3.1.1. QMLE: Homoskedastic Case

In this part, we establish consistency and asymptotic normality of the QMLE for the MESS(1,1) with i.i.d. disturbances.

Assumption 5. *The ϵ_{ni}'s are i.i.d. with mean zero and variance σ_0^2 and the moment $\mathrm{E}\,|\epsilon_{ni}|^{4+\eta}$ for some $\eta > 0$ exists.*

Define $\bar{Q}_n(\phi) = \min_\beta \mathrm{E}\, Q_n(\gamma)$, then

$$\begin{aligned} \bar{Q}_n(\phi) = {} & (X_n\beta_0)' e^{(\alpha-\alpha_0)W_n'} e^{\tau M_n'} H_n(\tau) e^{\tau M_n} e^{(\alpha-\alpha_0)W_n} X_n \beta_0 \\ & + \sigma_0^2 \operatorname{tr}(e^{-\tau_0 M_n'} e^{(\alpha-\alpha_0)W_n'} e^{\tau M_n'} e^{\tau M_n} e^{(\alpha-\alpha_0)W_n} e^{-\tau_0 M_n}). \end{aligned} \tag{10}$$

The identification of ϕ_0 can be based on the minimum values of $\{\bar{Q}_n(\phi)/n\}$. To ensure the identification uniqueness, the following condition is assumed.

[16] The main purpose for Assumption 3 is to guarantee that uniform convergence of relevant objects is possible on a compact parameter space.

Assumption 6. *Either (i)* $\lim_{n\to\infty} n^{-1}(X_n\beta_0)'e^{(\alpha-\alpha_0)W_n'}e^{\tau M_n'}H_n(\tau)e^{\tau M_n}e^{(\alpha-\alpha_0)W_n}X_n\beta_0 \neq 0$ *for any* τ *and* $\alpha \neq \alpha_0$*, and* $\lim_{n\to\infty} n^{-1}\operatorname{tr}(e^{(\tau-\tau_0)M_n'}e^{(\tau-\tau_0)M_n}) > 1$ *for any* $\tau \neq \tau_0$*, or*
(ii) $\lim_{n\to\infty} n^{-1}\operatorname{tr}(e^{-\tau_0 M_n'}e^{(\alpha-\alpha_0)W_n'}e^{\tau M_n'}e^{\tau M_n}e^{(\alpha-\alpha_0)W_n}e^{-\tau_0 M_n}) > 1$ *for any* $\phi \neq \phi_0$*.*

The identification of α_0 can come from the first term on the r.h.s. of (10). As $H_n(\tau)e^{\tau M_n}X_n = 0$, the first term at α_0 is zero for any τ. Thus the first term is not sufficient to identify τ_0. Given the identification of α_0, τ_0 can be identified from the second term. As $\lim_{n\to\infty} n^{-1}X_n'e^{\tau M_n'}e^{\tau M_n}X_n$ is non-singular, by the partition matrix formula, $\lim_{n\to\infty} n^{-1}(X_n\beta_0)'e^{(\alpha-\alpha_0)W_n'}e^{\tau M_n'}H_n(\tau)e^{\tau M_n}e^{(\alpha-\alpha_0)W_n}X_n\beta_0$ is non-zero if and only if $\lim_{n\to\infty} n^{-1}(X_n, e^{(\alpha-\alpha_0)W_n}X_n\beta_0)'e^{\tau M_n'}e^{\tau M_n}(X_n, e^{(\alpha-\alpha_0)W_n}X_n\beta_0)$ is nonsingular. Thus, the first part of (i) in Assumption 6 relates to asymptotic non-multicollinearity of $e^{(\alpha-\alpha_0)W_n}X_n\beta_0$ with X_n. In the proof of Proposition 1, it is shown by the inequality of arithmetic and geometric means that $n^{-1}\operatorname{tr}(e^{(\tau-\tau_0)M_n'}e^{(\tau-\tau_0)M_n}) \geq 1$ holds for any τ. The second part of (i) further requires $n^{-1}\operatorname{tr}(e^{(\tau-\tau_0)M_n'}e^{(\tau-\tau_0)M_n})$ to be strictly greater than 1 in the limit when $\tau \neq \tau_0$. For a finite n, the arithmetic and geometric means are equal if and only if all the eigenvalues of $e^{(\tau-\tau_0)M_n'}e^{(\tau-\tau_0)M_n}$ are equal to each other, which implies that $e^{(\tau-\tau_0)M_n'}e^{(\tau-\tau_0)M_n}$ is proportional to I_n. This assumption rules out this possibility in the limit whenever $\tau \neq \tau_0$. The identification of ϕ_0 can come only from the second term on the r.h.s. of (10), which is given in (ii) of Assumption 6. This relates to the uniqueness of the VC matrix of y_n, namely $\sigma_0^2 e^{-\alpha_0 W_n}e^{-\tau_0 M_n}e^{-\tau_0 M_n'}e^{-\alpha_0 W_n'}$, since $\operatorname{tr}(e^{-\tau_0 M_n'}e^{(\alpha-\alpha_0)W_n'}e^{\tau M_n'}e^{\tau M_n}e^{(\alpha-\alpha_0)W_n}e^{-\tau_0 M_n}) = \operatorname{tr}[e^{-\alpha_0 W_n}e^{-\tau_0 M_n}e^{-\tau_0 M_n'}e^{-\alpha_0 W_n'}(e^{-\alpha W_n}e^{-\tau M_n}e^{-\tau M_n'}e^{-\alpha W_n'})^{-1}]$.

It is obvious that Assumption 6 (i) fails to hold when $\beta_0 = 0$. In this case, the identification will rely solely on (ii). Another case of the failure of (i) even if $\beta_0 \neq 0$ occurs is when X_n contains only an intercept term, i.e., $X_n = l_n$, and W_n is row-normalized. In this case, $H_n(\tau)e^{\tau M_n}l_n = 0$. Other cases might be due to very special structures on W_n or M_n. For example, elements of W_n and M_n except the diagonal ones are all equal to a constant and X_n contains an intercept term. Let $W_n = M_n = (n-1)^{-1}(l_n l_n' - I_n)$ for instance. Then $H_n(\tau)e^{\tau M_n}W_n^k = (-1)^k(n-1)^{-k}H_n(\tau)e^{\tau M_n}$. By the expansion form of $e^{(\alpha-\alpha_0)W_n}$, $H_n(\tau)e^{\tau M_n}e^{(\alpha-\alpha_0)W_n}X_n = e^{(\alpha_0-\alpha)/(n-1)}H_n(\tau)e^{\tau M_n}X_n = 0$. Thus the first part in (i) does not hold. Furthermore, since the eigenvalues of M_n are $(1-n)^{-1}, \ldots, (1-n)^{-1}$ and 1, it follows that M_n^k has eigenvalues

$(1-n)^{-k}, \ldots, (1-n)^{-k}, 1$. Hence, with this symmetric M_n,

$$\begin{aligned}\frac{1}{n}\operatorname{tr}(e^{(\tau-\tau_0)M_n'}e^{(\tau-\tau_0)M_n}) &= \frac{1}{n}\operatorname{tr}(e^{2(\tau-\tau_0)M_n}) = \frac{1}{n}\sum_{k=0}^{\infty}\frac{2^k(\tau-\tau_0)^k\operatorname{tr}(M_n^k)}{k!} \\ &= \frac{1}{n}\sum_{k=0}^{\infty}\frac{2^k(\tau-\tau_0)^k[1+(n-1)(1-n)^{-k}]}{k!} \\ &= \frac{1}{n}e^{2(\tau-\tau_0)} + \frac{n-1}{n}e^{2(\tau-\tau_0)/(1-n)},\end{aligned}$$

which is equal to 1 in the limit. Then the second part in (i) does not hold either. In this case, $\lim_{n\to\infty}\frac{1}{n}\bar{Q}_n(\phi)$ is equal to σ_0^2 for any ϕ. Looking into $Q_n(\phi)$ directly, we have $Q_n(\phi) = e^{-2\alpha/(n-1)}y_n'e^{\tau M_n'}H_n(\tau)e^{\tau M_n}y_n$, which is monotonically decreasing in α. Then the QMLE of α will diverge to positive infinity, which is not equal to α_0.[17]

In general, (ii) in Assumption 6 would not hold as long as W_n and M_n are equal. When $M_n = W_n$,

$$\operatorname{tr}(e^{-\tau_0 M_n'}e^{(\alpha-\alpha_0)W_n'}e^{\tau M_n'}e^{\tau M_n}e^{(\alpha-\alpha_0)W_n}e^{-\tau_0 M_n}) = \operatorname{tr}(e^{(\alpha+\tau-\alpha_0-\tau_0)W_n'}e^{(\alpha+\tau-\alpha_0-\tau_0)W_n}).$$

As long as $\alpha+\tau = \alpha_0+\tau_0$, $\frac{1}{n}\operatorname{tr}(e^{-\tau_0 M_n'}e^{(\alpha-\alpha_0)W_n'}e^{\tau M_n'}e^{\tau M_n}e^{(\alpha-\alpha_0)W_n}e^{-\tau_0 M_n}) = 1$. So for the case that $M_n = W_n$, the parameter identification depends crucially on Assumption 6 (i). This situation is apparent as the model becomes $y_n = e^{-\alpha_0 W_n}X_n\beta_0 + e^{-(\alpha_0+\tau_0)W_n}\epsilon_n$. Thus, when there are no exogenous variables and $W_n = M_n$ in the MESS(1,1), α_0 and τ_0 cannot be identified.

With the identification uniqueness and uniform convergence of $[Q_n(\phi) - \bar{Q}_n(\phi)]/n$ to zero on the parameter space Φ, consistency of the QMLE follows.

Proposition 1. *Under Assumptions 1–6, the QMLE $\hat{\gamma}_n$ of the MESS(1,1) is consistent.*

The asymptotic distribution of $\hat{\gamma}_n$ can be derived from applying the mean value theorem to the first-order condition $\frac{\partial Q_n(\hat{\gamma}_n)}{\partial\gamma} = 0$ at the true γ_0, which yields $\sqrt{n}(\hat{\gamma}_n - \gamma_0) = -\left(\frac{1}{n}\frac{\partial^2 Q_n(\tilde{\gamma}_n)}{\partial\gamma\partial\gamma'}\right)^{-1}\frac{1}{\sqrt{n}}\frac{\partial Q_n(\gamma_0)}{\partial\gamma}$, where $\tilde{\gamma}_n$ is between $\hat{\gamma}_n$ and γ_0. Let $\mathbb{W}_n = e^{\tau_0 M_n}W_n e^{-\tau_0 M_n}$ and $A^s = A + A'$ for any square matrix A. Under some regularity conditions, $\frac{1}{n}\frac{\partial^2 Q_n(\tilde{\gamma}_n)}{\partial\gamma\partial\gamma'} = C_n + o_P(1)$, where

$$\begin{aligned}C_n &= \mathrm{E}\left(\frac{1}{n}\frac{\partial^2 Q_n(\gamma_0)}{\partial\gamma\partial\gamma'}\right) \\ &= \frac{1}{n}\begin{pmatrix}\sigma_0^2\operatorname{tr}(\mathbb{W}_n^s\mathbb{W}_n^s) + 2(\mathbb{W}_n e^{\tau_0 M_n}X_n\beta_0)'(\mathbb{W}_n e^{\tau_0 M_n}X_n\beta_0) & * & * \\ \sigma_0^2\operatorname{tr}(\mathbb{W}_n^s M_n^s) & \sigma_0^2\operatorname{tr}(M_n^s M_n^s) & * \\ -2(e^{\tau_0 M_n}X_n)'\mathbb{W}_n e^{\tau_0 M_n}X_n\beta_0 & 0 & 2(e^{\tau_0 M_n}X_n)'(e^{\tau_0 M_n}X_n)\end{pmatrix}.\end{aligned}$$

[17] See Smith (2009) for a discussion of this special case in the SAR model.

As $\operatorname{tr}(AB) = \operatorname{vec}'(A')\operatorname{vec}(B)$ for two conformable matrices A and B, where $\operatorname{vec}(\cdot)$ denotes the vectorization of a matrix, C_n may be written as $C_n = \frac{1}{n}C_{1n}'C_{1n}$, where

$$C_{1n} = \begin{pmatrix} \sigma_0 \operatorname{vec}(\mathbb{W}_n^s) & \sigma_0 \operatorname{vec}(M_n^s) & 0 \\ \sqrt{2}\mathbb{W}_n e^{\tau_0 M_n} X_n \beta_0 & 0 & -\sqrt{2} e^{\tau_0 M_n} X_n \end{pmatrix}. \tag{11}$$

Thus C_n is positive semi-definite. The following assumption guarantees that C_n is nonsingular in the limit.

Assumption 7. $\lim_{n\to\infty} \frac{1}{n}\Big((\mathbb{W}_n e^{\tau_0 M_n} X_n\beta_0)' H_n(\tau_0)(\mathbb{W}_n e^{\tau_0 M_n} X_n \beta_0) + \frac{\sigma_0^2}{2\operatorname{tr}(M_n^s M_n^s)}\big(\operatorname{tr}(\mathbb{W}_n^s\mathbb{W}_n^s)\operatorname{tr}(M_n^s M_n^s) - \operatorname{tr}^2(\mathbb{W}_n^s M_n^s)\big)\Big) \neq 0$ *and* $\lim_{n\to\infty}\frac{1}{n}\operatorname{tr}(M_n^s M_n^s) \neq 0$.

As $M_n^s M_n^s$ in the above assumption is positive semi-definite but not a zero matrix, $\operatorname{tr}(M_n^s M_n^s) > 0$. Note that $(\mathbb{W}_n e^{\tau_0 M_n} X_n\beta_0)' H_n(\tau_0)(\mathbb{W}_n e^{\tau_0 M_n} X_n\beta_0) \geq 0$, and $\operatorname{tr}(\mathbb{W}_n^s\mathbb{W}_n^s)\operatorname{tr}(M_n^s M_n^s) - \operatorname{tr}^2(\mathbb{W}_n^s M_n^s) \geq 0$ by the Cauchy-Schwarz inequality. By (8), the first-order derivatives of $Q_n(\gamma)$ at γ_0 have mean zero and are linear and quadratic functions of ϵ_n. Thus the central limit theorem for linear-quadratic forms in Kelejian and Prucha (2001) is applicable. Let $\mu_3 = \mathrm{E}\,\epsilon_{ni}^3$, $\mu_4 = \mathrm{E}\,\epsilon_{ni}^4$, and $\operatorname{vec}_{\mathrm{D}}(A)$ be a vector containing the diagonal elements of the square matrix A. The VC matrix Ω_n of $\frac{1}{\sqrt{n}}\frac{\partial Q_n(\gamma_0)}{\partial\gamma}$ is

$$\Omega_n = 2\sigma_0^2 C_n + \Omega_{1n} \text{ with } \Omega_{1n} = \frac{1}{n}\begin{pmatrix} (\mu_4 - 3\sigma_0^4)\operatorname{vec}_{\mathrm{D}}{}'(\mathbb{W}_n^s)\operatorname{vec}_{\mathrm{D}}(\mathbb{W}_n^s) + 4\mu_3(\mathbb{W}_n e^{\tau_0 M_n} X_n\beta_0)'\operatorname{vec}_{\mathrm{D}}(\mathbb{W}_n^s) & * & * \\ 0 & 0 & * \\ -2\mu_3(e^{\tau_0 M_n} X_n)'\operatorname{vec}_{\mathrm{D}}(\mathbb{W}_n^s) & 0 & 0 \end{pmatrix}.$$

When ϵ_{ni}'s are normal, $\mu_3 = \mu_4 - 3\sigma_0^4 = 0$; when $\tau_0 = 0$ or W_n and M_n are commutative, $\operatorname{vec}_{\mathrm{D}}(\mathbb{W}_n^s) = \operatorname{vec}_{\mathrm{D}}(W_n^s) = 0$ as W_n has a zero diagonal. These cases imply that $\Omega_{1n} = 0$ and Ω_n simplifies to $2\sigma_0^2 C_n$. As Ω_n is a VC matrix, it is positive semi-definite. We may also directly show that Ω_n is positive semi-definite. Note that $\mathrm{E}(\epsilon_{ni}^2 - \sigma_0^2)^2\,\mathrm{E}\,\epsilon_{ni}^2 \geq \big(\mathrm{E}[(\epsilon_{ni}^2 - \sigma_0^2)\epsilon_{ni}]\big)^2$, i.e. $(\mu_4 - \sigma_0^4)\sigma_0^2 \geq \mu_3^2$, by the Cauchy-Schwarz inequality. In addition, $\operatorname{tr}(\operatorname{Diag}(\mathbb{W}_n^s)\operatorname{Diag}(\mathbb{W}_n^s)) = \operatorname{vec}_{\mathrm{D}}{}'(\mathbb{W}_n^s)\operatorname{vec}_{\mathrm{D}}(\mathbb{W}_n^s)$, and $\operatorname{tr}(\operatorname{Diag}(\mathbb{W}_n^s)M_n^s) = 0$ as M_n has a zero diagonal, where $\operatorname{Diag}(A)$ for a square matrix A denotes a diagonal matrix whose diagonal is equal to that of A. Then Ω_n can be written as $\Omega_n = \frac{1}{n}\Omega_{2n}'\Omega_{2n}$, where

$$\Omega_{2n} = \begin{pmatrix} \sqrt{2}\sigma_0^2 \operatorname{vec}\big(\mathbb{W}_n^s - \operatorname{Diag}(\mathbb{W}_n^s) + \sqrt{\frac{\sigma_0^2(\mu_4-\sigma_0^4)-\mu_3^2}{2\sigma_0^6}}\operatorname{Diag}(\mathbb{W}_n^s)\big) & \sqrt{2}\sigma_0^2\operatorname{vec}(M_n^s) & 0 \\ 2\sigma_0\mathbb{W}_n e^{\tau_0 M_n} X_n\beta_0 + \frac{\mu_3}{\sigma_0}\operatorname{vec}_{\mathrm{D}}(\mathbb{W}_n^s) & 0 & -2\sigma_0 e^{\tau_0 M_n} X_n \end{pmatrix}.$$

Thus Ω_n is positive semi-definite.

Proposition 2. *Under Assumptions 1–7,* $\sqrt{n}(\hat{\gamma}_n - \gamma_0) \xrightarrow{d} N(0, \lim_{n\to\infty} C_n^{-1}\Omega_n C_n^{-1})$. *If* $\epsilon_n \sim N(0, \sigma_0^2 I_n)$*;* $\tau_0 = 0$*; or* W_n *and* M_n *are commutative, then* $\sqrt{n}(\hat{\gamma}_n - \gamma_0) \xrightarrow{d} N(0, 2\sigma_0^2 \lim_{n\to\infty} C_n^{-1})$.

When the disturbances ϵ_{ni}'s are normal, the generalized information matrix equality holds, thus the limiting distribution of the MLE $\hat{\gamma}_n$ does not depend on moments of the disturbances higher than the second order. Even when the disturbances ϵ_{ni}'s are not normally distributed, if there is no MESS process in the disturbances or the spatial weights matrices M_n and W_n are commutative, the limiting distribution of the QMLE does not involve moments of the disturbances higher than the second order.

3.1.2. QMLE: Heteroskedastic Case when W_n and M_n are Commutative

When the disturbances ϵ_{ni}'s are independent but may have different variances, the following assumption is made about the disturbances.

Assumption 8. *The ϵ_{ni}'s are independent $(0, \sigma_{ni}^2)$ and the moments $\mathrm{E}\,|\epsilon_{ni}|^{4+\eta}$ for some $\eta > 0$ exist and are uniformly bounded for all n and i.*

As argued earlier, when W_n and M_n can commute, or $\tau_0 = 0$, the minimization of the function $Q_n(\gamma)$ may yield a consistent estimator $\hat{\gamma}_n$ of γ under unknown heteroskedasticity, since the first-order derivatives of $Q_n(\gamma)$ at γ_0 have zero expectation. In practice for some situations, one may use a single spatial weights matrix W_n for both the main equation and the disturbance process, which implies the commutative property.

Assumption 9. *W_n and M_n are commutative or $\tau_0 = 0$.*

Define $\bar{Q}_n(\phi) = \min_\beta \mathrm{E}\, Q_n(\gamma)$. The identification of ϕ_0 can be based on minimizers of $\{\bar{Q}_n(\phi)\}$. Using Assumption 9, we have $\bar{Q}_n(\phi) = \bar{Q}_{1n}(\phi) + \bar{Q}_{2n}(\phi)$, where $\bar{Q}_{1n}(\phi) = (X_n\beta_0)' e^{(\alpha-\alpha_0)W_n'} e^{\tau M_n'} H_n(\tau) e^{\tau M_n} e^{(\alpha-\alpha_0)W_n} X_n\beta_0$ and $\bar{Q}_{2n}(\phi) = \mathrm{tr}(e^{(\alpha-\alpha_0)W_n'} e^{(\tau-\tau_0)M_n'} e^{(\tau-\tau_0)M_n} e^{(\alpha-\alpha_0)W_n} \Sigma_n)$. It is obvious that $\bar{Q}_{1n}(\phi) \geq 0$ and $\bar{Q}_{1n}(\phi_0) = 0$. As W_n and M_n have zero diagonals and Σ_n is a diagonal matrix, $\frac{\partial \bar{Q}_{2n}(\phi_0)}{\partial \alpha} = \mathrm{tr}[(W_n' + W_n)\Sigma_n] = 0$ and $\frac{\partial \bar{Q}_{2n}(\phi_0)}{\partial \tau} = \mathrm{tr}[(M_n' + M_n)\Sigma_n] = 0$. Thus ϕ_0 is a stationary point of $\bar{Q}_{2n}(\phi)$ and also $\bar{Q}_n(\phi)$. Using the commutative property of W_n and M_n, we have $\frac{\partial^2 \bar{Q}_{2n}(\phi)}{\partial \alpha^2} = \mathrm{tr}[\Sigma_n^{1/2} e^{(\alpha-\alpha_0)W_n'} e^{(\tau-\tau_0)M_n'} (W_n'^2 + W_n^2 + 2W_n'W_n) e^{(\tau-\tau_0)M_n} e^{(\alpha-\alpha_0)W_n} \Sigma_n^{1/2}]$,

$$\frac{\partial^2 \bar{Q}_{2n}(\phi)}{\partial \tau^2} = \mathrm{tr}[\Sigma_n^{1/2} e^{(\alpha-\alpha_0)W_n'} e^{(\tau-\tau_0)M_n'} (M_n'^2 + M_n^2 + 2M_n'M_n) e^{(\tau-\tau_0)M_n} e^{(\alpha-\alpha_0)W_n} \Sigma_n^{1/2}]$$

and $\frac{\partial^2 \bar{Q}_{2n}(\phi)}{\partial \alpha \partial \tau} = \mathrm{tr}\{\Sigma_n^{1/2} e^{(\alpha-\alpha_0)W_n'} e^{(\tau-\tau_0)M_n'} [(W_n' + W_n)M_n + M_n'W_n + W_n'M_n'] e^{(\tau-\tau_0)M_n} e^{(\alpha-\alpha_0)W_n} \Sigma_n^{1/2}\}$. If $W_n'W_n = W_nW_n'$, then $W_n'^2 + W_n^2 + 2W_n'W_n = (W_n' + W_n)^2$; if $M_n'M_n = M_nM_n'$, then $M_n'^2 + M_n^2 + 2M_n'M_n = (M_n' + M_n)^2$; if $M_n'W_n = W_nM_n'$, then $(W_n' + W_n)M_n + M_n'W_n + W_n'M_n' = (W_n' + W_n)(M_n' + M_n)$. Thus, under the conditions that $W_n'W_n = W_nW_n'$, $M_n'M_n = M_nM_n'$ and $M_n'W_n = W_nM_n'$, by the Cauchy-Schwarz inequality, $\frac{\partial^2 \bar{Q}_{2n}(\phi)}{\partial \alpha^2} \frac{\partial^2 \bar{Q}_{2n}(\phi)}{\partial \tau^2} \geq \left(\frac{\partial^2 \bar{Q}_{2n}(\phi)}{\partial \alpha \partial \tau}\right)^2$. In this case, $\frac{\partial^2 \bar{Q}_{2n}(\phi)}{\partial \phi \partial \phi'}$ is positive semi-definite and $\bar{Q}_{2n}(\phi)$ is a

concave function. It follows that ϕ_0 is a global minimizer of $\bar{Q}_{2n}(\phi)$ and $\bar{Q}_n(\phi)$. Hence, with some extra conditions on W_n and M_n, it is possible that ϕ_0 uniquely minimizes $\bar{Q}_n(\phi)/n$ in the limit.

It is also possible that ϕ_0 is only a local minimizer of $\bar{Q}_n(\phi)$. For example, in the case that $W_n = M_n$, $\frac{\partial^2 \bar{Q}_{2n}(\phi_0)}{\partial \alpha^2} = \frac{\partial^2 \bar{Q}_{2n}(\phi_0)}{\partial \tau^2} = \frac{\partial^2 \bar{Q}_{2n}(\phi_0)}{\partial \alpha \partial \tau} = \mathrm{tr}[(W_n'^2 + W_n^2 + 2W_n' W_n)\Sigma_n]$, which is positive if elements of W_n are non-negative. Then $\frac{\partial^2 \bar{Q}_{2n}(\phi_0)}{\partial \phi \partial \phi'}$ is positive semi-definite and $\bar{Q}_{2n}(\phi)$ is concave at ϕ_0. Hence, ϕ_0 is a local minimizer of $\bar{Q}_{2n}(\phi)$ and $\bar{Q}_n(\phi)$. These considerations motivate the following identification condition.

Assumption 10. $\lim_{n\to\infty} \frac{1}{n}[\bar{Q}_n(\phi) - \mathrm{tr}(\Sigma_n)] > 0$ *for any* $\phi \neq \phi_0$.

Proposition 3. *Under Assumptions 1–4 and 8–10, the QMLE* $\hat{\gamma}_n$ *is consistent for* γ_0.

Let $D_n = \frac{1}{n}\,\mathrm{E}\big(\frac{\partial^2 Q_n(\gamma_0)}{\partial \gamma \partial \gamma'}\big)$ and $\Delta_n = \frac{1}{n}\,\mathrm{E}\big(\frac{\partial Q_n(\gamma_0)}{\partial \gamma}\frac{\partial Q_n(\gamma_0)}{\partial \gamma'}\big)$, then

$$D_n = \frac{2}{n}\begin{pmatrix} \mathrm{tr}(W_n^s W_n \Sigma_n) + (W_n e^{\tau_0 M_n} X_n \beta_0)'(W_n e^{\tau_0 M_n} X_n \beta_0) & * & * \\ \mathrm{tr}(M_n^s W_n \Sigma_n) & \mathrm{tr}(M_n^s M_n \Sigma_n) & * \\ -(e^{\tau_0 M_n} X_n)' W_n e^{\tau_0 M_n} X_n \beta_0 & 0 & (e^{\tau_0 M_n} X_n)'(e^{\tau_0 M_n} X_n) \end{pmatrix},$$

and

$$\Delta_n = \frac{2}{n}\begin{pmatrix} \mathrm{tr}(\Sigma_n W_n^s \Sigma_n W_n^s) + 2(W_n e^{\tau_0 M_n} X_n \beta_0)' \Sigma_n (W_n e^{\tau_0 M_n} X_n \beta_0) & * & * \\ \mathrm{tr}(\Sigma_n M_n^s \Sigma_n W_n^s) & \mathrm{tr}(\Sigma_n M_n^s \Sigma_n M_n^s) & * \\ -2(e^{\tau_0 M_n} X_n)' \Sigma_n W_n e^{\tau_0 M_n} X_n \beta_0 & 0 & 2(e^{\tau_0 M_n} X_n)' \Sigma_n (e^{\tau_0 M_n} X_n) \end{pmatrix}.$$

Note that Δ_n, being the VC matrix of a vector of linear-quadratic forms of disturbances, does not involve higher than the second moments of disturbances, because W_n and M_n in the quadratic forms $\epsilon_n' W_n^s \epsilon_n$ and $\epsilon_n' M_n^s \epsilon_n$ have zero diagonals (see Lee, 2007). We may write Δ_n as $\Delta_n = \frac{1}{n}\Delta_{1n}'\Delta_{1n}$, where

$$\Delta_{1n} = \begin{pmatrix} \sqrt{2}\,\mathrm{vec}(\Sigma_n^{1/2} W_n^s \Sigma_n^{1/2}) & \sqrt{2}\,\mathrm{vec}(\Sigma_n^{1/2} M_n^s \Sigma_n^{1/2}) & 0 \\ 2\Sigma_n^{1/2} W_n e^{\tau_0 M_n} X_n \beta_0 & 0 & -2\Sigma_n^{1/2} e^{\tau_0 M_n} X_n \end{pmatrix},$$

thus Δ_n is positive semi-definite. To make sure that D_n is invertible for large enough n, we need the following assumption.

Assumption 11. $\lim_{n\to\infty} \frac{1}{n}\,\mathrm{tr}(M_n^s M_n \Sigma_n) \neq 0$ *and* $\lim_{n\to\infty} \frac{1}{n}\big((W_n e^{\tau_0 M_n} X_n \beta_0)' H_n(\tau_0)(W_n e^{\tau_0 M_n} X_n \beta_0) + \frac{\mathrm{tr}(W_n^s W_n \Sigma_n)\,\mathrm{tr}(M_n^s M_n \Sigma_n) - \mathrm{tr}^2(M_n^s W_n \Sigma_n)}{\mathrm{tr}(M_n^s M_n \Sigma_n)}\big) \neq 0$.

When elements of W_n and M_n are non-negative, $\mathrm{tr}(M_n^s M_n \Sigma_n) > 0$, $\mathrm{tr}(M_n^s W_n \Sigma_n) \geq 0$ and $\mathrm{tr}(W_n^s W_n \Sigma_n) > 0$, because M_n and W_n are not zero matrices and the diagonal elements of Σ_n are positive in general.

Proposition 4. *Under Assumptions 1–4 and 8–11,* $\sqrt{n}(\hat{\gamma}_n - \gamma_0) \xrightarrow{d} N(0, \lim_{n\to\infty} D_n^{-1}\Delta_n D_n^{-1})$.

With the requirement of $\tau_0 = 0$ or commutativeness of W_n and M_n, in addition to the consistency, the QMLE under unknown heteroskedasticity has an asymptotic distribution that does not involve higher than the second moments of the disturbances, whether the disturbances are normal or not.

To make asymptotically valid inference using the QMLE $\hat{\gamma}_n$ under unknown heteroskedasticity, we need a consistent estimator for $D_n^{-1}\Delta_n D_n^{-1}$. As in White (1980), we may have a consistent estimator of $D_n^{-1}\Delta_n D_n^{-1}$ without being able to consistently estimate Σ_n, which has n unknown parameters. Let $\hat{\Sigma}_n = \mathrm{Diag}(\hat{\epsilon}_{n1}^2, \ldots, \hat{\epsilon}_{nn}^2)$, where $\hat{\epsilon}_n = (\hat{\epsilon}_{n1}, \ldots, \hat{\epsilon}_{nn})'$ is the residual vector from the QML estimation. Consistent estimators for D_n and Δ_n can be, respectively, $\hat{D}_n$ and $\hat{\Delta}_n$, which are the matrices derived from replacing Σ_n in D_n and Ω_n by $\hat{\Sigma}_n$, and replacing γ_0 by a consistent estimator $\hat{\gamma}_n$. The $\hat{D}_n$ and $\hat{\Delta}_n$ can be consistent because D_n and Δ_n with fixed dimensions are estimated as whole terms.

Proposition 5. *Under Assumptions 1–4 and 8–10, $\hat{D}_n = D_n + o_P(1)$ and $\hat{\Delta}_n = \Delta_n + o_P(1)$.*

3.2. GMME

We now consider the GMM estimation of the MESS(1,1). Let the moment vector be

$$g_n(\gamma) = \frac{1}{n}\big(\epsilon_n'(\gamma)P_{n1}\epsilon_n(\gamma), \ldots, \epsilon_n'(\gamma)P_{n,k_p}\epsilon_n(\gamma), \epsilon_n'(\gamma)F_n\big)', \tag{12}$$

where $\epsilon_n(\gamma) = e^{\tau M_n}(e^{\alpha W_n}y_n - X_n\beta)$, the n-dimensional square matrices P_{ni}'s for the quadratic moments have zero traces when ϵ_{ni}'s are i.i.d. and have zero diagonals when ϵ_{ni}'s are independent but with different variances, and the $n \times k_f$ instrumental variable matrix F_n used in the 2SLS approach can consist of the independent columns of X_n, W_nX_n, M_nX_n, $W_n^2X_n$, $M_n^2X_n$ and so on.[18] The GMM objective function with the weighting matrix a_na_n' is $g_n'(\gamma)a_na_n'g_n(\gamma)$, where the full column rank $(k_p + k_f) \times k_a$ matrix a_n with $k_a \geq k + 2$ converges to a full rank matrix a_0 by design.

3.2.1. GMME: Homoskedastic Case

When the disturbances are i.i.d., the GMME can be consistent when the matrices P_{ni}'s have zero traces but not necessarily zero diagonals. The P_{ni}'s are constructed from W_n and M_n, thus we may assume that P_{ni}'s are bounded in row and column sum norms.

Assumption 12. *The n-dimensional square matrices $P_{n1}, \ldots, P_{n,k_p}$ have zero traces and are bounded in both row and column sum norms. Elements of F_n are uniformly bounded constants.*

[18] For α and β, we may use only the linear instrument F_n and implement a 2SLS estimation, for which the objective function is $(e^{\alpha W_n}y_n - X_n\beta)'F_n(F_n'F_n)^{-1}F_n'(e^{\alpha W_n}y_n - X_n\beta)$ or $(e^{\alpha W_n}y_n - X_n\beta)'e^{\hat{\tau}_n M_n'}F_n(F_n'F_n)^{-1}F_n'e^{\hat{\tau}_n M_n}(e^{\alpha W_n}y_n - X_n\beta)$ when taking into account the MESS process in the disturbances, where $\hat{\tau}_n$ is an initial consistent estimator of τ. This is a nonlinear 2SLS that does not have a closed-form solution. Thus it does not have a computational advantage as the traditional 2SLS and we do not discuss it separately.

For any γ,

$$\begin{aligned}\mathrm{E}[\epsilon_n'(\gamma)P_{ni}\epsilon_n(\gamma)] &= (e^{(\alpha-\alpha_0)W_n}X_n\beta_0 - X_n\beta)'e^{\tau M_n'}P_{ni}e^{\tau M_n}(e^{(\alpha-\alpha_0)W_n}X_n\beta_0 - X_n\beta) \\ &\quad + \sigma_0^2\,\mathrm{tr}(e^{-\tau_0 M_n'}e^{(\alpha-\alpha_0)W_n'}e^{\tau M_n'}P_{ni}e^{\tau M_n}e^{(\alpha-\alpha_0)W_n}e^{-\tau_0 M_n}),\end{aligned} \tag{13}$$

$$\mathrm{E}[F_n'\epsilon_n(\gamma)] = F_n'e^{\tau M_n}(e^{(\alpha-\alpha_0)W_n}X_n\beta_0 - X_n\beta). \tag{14}$$

The identification of γ_0 requires a unique solution of the limiting equations $\lim_{n\to\infty}\mathrm{E}\,g_n(\gamma) = 0$ at γ_0. When $\alpha = \alpha_0$ and $\beta = \beta_0$, $\mathrm{E}[F_n'\epsilon_n(\gamma)] = 0$ whatever τ is. Thus τ cannot be identified from the linear moments $\mathrm{E}[F_n'\epsilon_n(\gamma)] = 0$, because it only plays a role as weighting. It is possible that α_0 and β_0 may be identified from $\mathrm{E}[F_n'\epsilon_n(\gamma)] = 0$, and τ_0 be identified from the quadratic moments $\mathrm{E}[\epsilon_n'(\gamma)P_{ni}\epsilon_n(\gamma)] = 0$, $i = 1,\ldots,k_p$. Let $F_n = (F_{1n}, F_{2n})$ such that $\lim_{n\to\infty}\frac{1}{n}F_{2n}'e^{\tau M_n}X_n$ is nonsingular for any $\tau \in [-\delta,\delta]$, which is a part of a rank condition for valid IV's. The $\mathrm{E}[F_n'\epsilon_n(\gamma)] = 0$ is equivalent to $F_{1n}'e^{\tau M_n}(e^{(\alpha-\alpha_0)W_n}X_n\beta_0 - X_n\beta) = 0$ and $F_{2n}'e^{\tau M_n}(e^{(\alpha-\alpha_0)W_n}X_n\beta_0 - X_n\beta) = 0$. From the equation involving only F_{2n}, we have $\beta = (F_{2n}'e^{\tau M_n}X_n)^{-1}F_{2n}'e^{\tau M_n}e^{(\alpha-\alpha_0)W_n}X_n\beta_0$. With substitution, the equation involving F_{1n} becomes $F_{1n}'H_{1n}(\tau)e^{\tau M_n}e^{(\alpha-\alpha_0)W_n}X_n\beta_0 = 0$, where $H_{1n}(\tau) = I_n - e^{\tau M_n}X_n(F_{2n}'e^{\tau M_n}X_n)^{-1}F_{2n}'$. Furthermore, it reduces to $F_{1n}'H_{1n}(\tau)e^{\tau M_n}X_n\beta_0 = 0$ when $\alpha = \alpha_0$. In the case that α_0 can be identified from the equation, $F_{1n}'H_{1n}(\tau)e^{\tau M_n}e^{\eta W_n}X_n\beta_0 \neq 0$ for any $\eta \neq 0$. When $\alpha = \alpha_0$ and $\beta = \beta_0$, (13) becomes $\sigma_0^2\,\mathrm{tr}(e^{(\tau-\tau_0)M_n'}P_{ni}e^{(\tau-\tau_0)M_n}) = 0$. Then the identification of τ_0 requires some matrix P_{ni} such that $\lim_{n\to\infty}\frac{1}{n}\,\mathrm{tr}(e^{(\tau-\tau_0)M_n'}P_{ni}e^{(\tau-\tau_0)M_n}) \neq 0$ when $\tau \neq \tau_0$. It is also possible that α_0 cannot be identified from the linear moment (14), then the identification of (α_0,τ_0) would be from the quadratic moments.[19]

Assumption 13. *Suppose that F_n may be written as $F_n = (F_{1n}, F_{2n})$ such that $\lim_{n\to\infty}\frac{1}{n}F_{2n}'e^{\tau M_n}X_n$ is nonsingular for any $\tau \in [-\delta,\delta]$. Furthermore, either 1) $\lim_{n\to\infty}\frac{1}{n}F_{1n}'H_{1n}(\tau)e^{\tau M_n}e^{\eta W_n}X_n\beta_0 \neq 0$ for any $\eta \neq 0$ and for all $\tau \in [-\delta,\delta]$; and, for any $\tau \neq \tau_0$, $\lim_{n\to\infty}\frac{1}{n}\mathrm{tr}(e^{(\tau-\tau_0)M_n'}P_{ni}e^{(\tau-\tau_0)M_n}) \neq 0$, for some $i \in \{1,\cdots,k_p\}$; or, 2) for any $(\alpha,\tau) \neq (\alpha_0,\tau_0)$, $\lim_{n\to\infty}\frac{1}{n}\beta_0'X_n'e^{(\alpha-\alpha_0)W_n'}e^{\tau M_n'}H_{1n}'(\tau)P_{ni}H_{1n}(\tau)e^{\tau M_n}e^{(\alpha-\alpha_0)W_n}X_n\beta_0 + \frac{\sigma_0^2}{n}\,\mathrm{tr}(e^{-\tau_0 M_n'}e^{(\alpha-\alpha_0)W_n'}e^{\tau M_n'}P_{ni}e^{\tau M_n}e^{(\alpha-\alpha_0)W_n}e^{-\tau_0 M_n}) \neq 0$, for some $i \in \{1,\cdots,k_p\}$.*

As usual for nonlinear extremum estimators, we assume the compactness of the parameter space of γ (Amemiya, 1985).

Assumption 14. *The parameter space Γ of γ is compact and the true γ_0 is in the interior of Γ.*

Proposition 6. *Under Assumptions 1, 2, 5 and 12–14, the GMM estimator $\hat{\gamma}_n$ from the minimization of $g_n'(\gamma)a_na_n'g_n(\gamma)$ is a consistent estimator of γ_0, and*

$$\sqrt{n}(\hat{\gamma}_n - \gamma_0) \xrightarrow{d} N\big(0, \lim_{n\to\infty}(G_n'a_na_n'G_n)^{-1}G_n'a_na_n'V_na_na_n'G_n(G_n'a_na_n'G_n)^{-1}\big),$$

[19] For example, this can occur when F_{1n} is linearly dependent on F_{2n}.

where $V_n = n\,\mathrm{E}[g_n(\gamma_0)g_n'(\gamma_0)] = \frac{1}{n}\begin{pmatrix} \frac{\sigma_0^4}{2}\omega_n'\omega_n + \frac{1}{4}(\mu_4 - 3\sigma_0^4)\omega_{nd}'\omega_{nd} & \frac{1}{2}\mu_3\omega_{nd}'F_n \\ \frac{1}{2}\mu_3 F_n'\omega_{nd} & \sigma_0^2 F_n'F_n \end{pmatrix}$ *and*

$$G_n = \mathrm{E}\,\frac{\partial g_n(\gamma_0)}{\partial \gamma'} = \frac{1}{n}\begin{pmatrix} \frac{\sigma_0^2}{2}\omega_n'\,\mathrm{vec}(\mathbb{W}_n^s) & \frac{\sigma_0^2}{2}\omega_n'\,\mathrm{vec}(M_n^s) & 0 \\ F_n'\mathbb{W}_n e^{\tau_0 M_n}X_n\beta_0 & 0 & -F_n'e^{\tau_0 M_n}X_n \end{pmatrix},$$

with $\omega_n = \big(\mathrm{vec}(P_{n1}^s),\ldots,\mathrm{vec}(P_{n,k_p}^s)\big)$ *and* $\omega_{nd} = \big(\mathrm{vec_D}(P_{n1}^s),\ldots,\mathrm{vec_D}(P_{n,k_p}^s)\big)$*, under the condition that* $\lim_{n\to\infty} a_n'G_n$ *exists and has the full rank* $k+2$*.*

Within the GMM framework, with moments $g_n(\gamma)$, an optimum GMM will use V_n^{-1} as the optimum weighting in place of $a_n a_n'$. The variance matrix V_n of $g_n(\gamma_0)$ in the preceding proposition can be put into a more informative form as a positive semi-definite matrix. Let $\omega_n^{\#} = (\mathrm{vec}(P_{n1}^{\#s}),\ldots,\mathrm{vec}(P_{n,k_p}^{\#s}))$, where $P_{ni}^{\#s} = \frac{1}{2}\sqrt{\mu_4 - \sigma_0^4 - \frac{\mu_3^2}{\sigma_0^2}}\,\mathrm{Diag}(P_{ni}^s) + \frac{\sqrt{2}\sigma_0^2}{2}[P_{ni}^s - \mathrm{Diag}(P_{ni}^s)]$, then $V_n = \frac{1}{n}\begin{pmatrix} \omega_n^{\#} & 0 \\ \frac{\mu_3}{2\sigma_0}\omega_{nd} & \sigma_0 F_n \end{pmatrix}'\begin{pmatrix} \omega_n^{\#} & 0 \\ \frac{\mu_3}{2\sigma_0}\omega_{nd} & \sigma_0 F_n \end{pmatrix}$. Thus V_n is positive semi-definite. We require the non-singularity of V_n to formulate the feasible optimal GMM, which is guaranteed by the following assumption.

Assumption 15. *The limits of* $\frac{1}{n}F_n'F_n$ *and* $\frac{\sigma_0^4}{2n}(\omega_n'\omega_n - \omega_{nd}'\omega_{nd}) + \frac{1}{4n}(\mu_4 - \sigma_0^2 - \frac{\mu_3^2}{\sigma_0^2})\omega_{nd}'\omega_{nd} + \frac{\mu_3^2}{4n\sigma_0^2}\omega_{nd}'H_{F_n}\omega_{nd}$ *exist and are nonsingular, where* $H_{F_n} = I_n - F_n(F_n'F_n)^{-1}F_n'$*.*

Note that $\omega_n'\omega_n - \omega_{nd}'\omega_{nd} = \big(\mathrm{vec}(P_{n1}^{\#s} - \mathrm{Diag}(P_{n1}^{\#s})),\ldots,\mathrm{vec}(P_{n,k_p}^{\#s} - \mathrm{Diag}(P_{n,k_p}^{\#s}))\big)'\big(\mathrm{vec}(P_{n1}^{\#s} - \mathrm{Diag}(P_{n1}^{\#s})),\ldots,\mathrm{vec}(P_{n,k_p}^{\#s} - \mathrm{Diag}(P_{n,k_p}^{\#s}))\big) \geq 0$. When $\lim_{n\to\infty}\frac{1}{n}F_n'F_n$ is nonsingular, the above assumption is satisfied as long as one of the terms $\lim_{n\to\infty}\frac{1}{n}(\omega_n'\omega_n - \omega_{nd}'\omega_{nd})$, $\lim_{n\to\infty}\frac{1}{n}\omega_{nd}'\omega_{nd}$, and $\lim_{n\to\infty}\frac{\mu_3^2}{n}\omega_{nd}'H_{F_n}\omega_{nd}$ is nonsingular. A consistent estimator $\hat{V}_n$ for V_n may be obtained from replacing the σ_0^2, μ_3 and μ_4 in V_n by their consistent estimators.

Proposition 7. *Under Assumptions 1, 2, 5 and 12–15, the feasible optimal GMME* $\hat{\gamma}_{n,o}$ *from the minimization of* $g_n'(\gamma)\hat{V}_n^{-1}g_n(\gamma)$ *is a consistent estimator of* γ_0*, and* $\sqrt{n}(\hat{\gamma}_{n,o} - \gamma_0) \xrightarrow{d} N\big(0, \lim_{n\to\infty}(G_n'V_n^{-1}G_n)^{-1}\big)$*.*

As the possible selections of linear and quadratic moments via F_n and P_{ni}'s are numerous, there is an issue regarding the best design for those matrices. For that purpose, we follow Breusch et al. (1999) to show that additional linear and quadratic moments are redundant given properly selected ones.[20] If $e^{\tau_0 M_n}X_n$ contains an intercept term due to the presence of an intercept term in X_n, let X_n^* be the submatrix of X_n with the intercept term deleted, so that $e^{\tau_0 M_n}X_n = [e^{\tau_0 M_n}X_n^*, c(\tau_0)l_n]$, where $c(\tau_0)$ is a scalar function of

[20] This pursuit is motivated by that in Liu et al. (2010).

τ_0.[21] Otherwise, $X_n^* = X_n$ and $e^{\tau_0 M_n} X_n^* = e^{\tau_0 M_n} X_n$. Suppose that there are k^* columns in X_n^*. Let X_{nl}^* be the lth column of X_n^*, $\eta_3 = \mu_3 \sigma_0^{-3}$ and $\eta_4 = \mu_4 \sigma_0^{-4}$ be the skewness and kurtosis of the disturbances. Furthermore, let $A_n^{(t)} = A_n - I_n \operatorname{tr}(A_n)/n$ for any $n \times n$ matrix A_n, which is the matrix A_n with its trace subtracted out from its diagonal. Thus $A_n^{(t)}$ has zero trace. The following proposition gives the moment conditions for the GMME that generate the smallest variance within the class of all GMM estimators with linear and quadratic moments, when disturbances are homoskedastic.

Proposition 8. *Suppose that Assumptions 1, 2, 5, and 12–15 hold. Let* $g_n^*(\gamma) = \frac{1}{n}(P_{n1}^* \epsilon(\gamma), \ldots, P_{n,k^*+4}^* \epsilon(\gamma), F_n^{*\prime} \epsilon(\gamma)$, *where* $P_{n1}^* = \mathbb{W}_n$, $P_{n2}^* = \operatorname{Diag}(\mathbb{W}_n)$, $P_{n3}^* = \operatorname{Diag}(e^{\tau_0 M_n} W_n X_n \beta_0)^{(t)}$, $P_{n4}^* = M_n$, $P_{n,l+4}^* = \operatorname{Diag}(e^{\tau_0 M_n} X_{nl}^*)^{(t)}$ *for* $l = 1, \ldots, k^*$, *and* $F_n^* = (F_{n1}^*, F_{n2}^*, F_{n3}^*, F_{n4}^*)$ *with* $F_{n1}^* = e^{\tau_0 M_n} X_n^*$, $F_{n2}^* = e^{\tau_0 M_n} W_n X_n \beta_0$, $F_{n3}^* = l_n$, *and* $F_{n4}^* = \operatorname{vec_D}(\mathbb{W}_n)$. *Denote* $V_n^* = n \operatorname{E}[g_n^*(\gamma_0) g_n^{*\prime}(\gamma_0)]$. *Then* $\hat{\gamma}_n^* = \min_\gamma g_n^{*\prime}(\gamma) V_n^{*-1} g_n^*(\gamma)$ *is the best GMME within the class of GMMEs with linear and quadratic moments, and* $\hat{\gamma}_n^*$ *has the asymptotic distribution that* $\sqrt{n}(\hat{\gamma}_n^* - \gamma_0) \xrightarrow{d} N\big(0, \lim_{n\to\infty} \Lambda_n^{*-1}\big)$, *where* $\Lambda_n^* = G_n^{*\prime} V_n^{*-1} G_n^*$ *with* $G_n^* = \operatorname{E} \frac{\partial g_n^*(\gamma_0)}{\partial \gamma'}$.

The detailed proof of this proposition is in Appendix A. From the proof, Λ_n^* has the following expression

$$\Lambda_n^* = \frac{1}{n} \begin{pmatrix} \operatorname{tr}(P_{\alpha n}^{*s} \mathbb{W}_n) + \sigma_0^{-2} (e^{\tau_0 M_n} W_n X_n \beta_0)' F_{\alpha n}^* & * & * \\ \operatorname{tr}(P_{\tau n}^{*s} \mathbb{W}_n) & \operatorname{tr}(P_{\tau n}^{*s} M_n) & * \\ -\sigma_0^{-2} (e^{\tau_0 M_n} X_n)' F_{\alpha n}^* & 0 & \sigma_0^{-2} (e^{\tau_0 M_n} X_n)' F_{\beta n}^* \end{pmatrix}, \tag{15}$$

where $P_{\alpha n}^* = P_{n1}^* - \frac{(\eta_4-3)-\eta_3^2}{(\eta_4-1)-\eta_3^2} P_{n2}^* - \frac{\sigma_0^{-1}\eta_3}{(\eta_4-1)-\eta_3^2} P_{n3}^*$, $P_{\tau n}^* = M_n$, $P_{\beta nl}^* = P_{n,l+4}^*$ for $l = 1, \ldots, k^*$, $F_{\alpha n}^* = \frac{\eta_4-1}{(\eta_4-1)-\eta_3^2} F_{n2}^* - \frac{\eta_3^2}{(\eta_4-1)-\eta_3^2} F_{n3}^* (\frac{1}{n} l_n' e^{\tau_0 M_n} W_n X_n \beta_0) - \frac{2\sigma_0 \eta_3}{(\eta_4-1)-\eta_3^2} F_{n4}^*$,

$$F_{\beta n}^* = \frac{\eta_4 - 1}{(\eta_4 - 1) - \eta_3^2} F_{n1}^* - \frac{\eta_3^2}{(\eta_4 - 1) - \eta_3^2} F_{n3}^* (\frac{1}{n} l_n' e^{\tau_0 M_n} X_n^*)$$

if $e^{\tau_0 M_n} X_n$ does not contain an intercept term; otherwise

$$F_{\beta n}^* = \frac{\eta_4 - 1}{(\eta_4 - 1) - \eta_3^2} F_{n1}^* (I_{k^*}, 0_{k^* \times 1}) + \frac{\eta_4 - 1}{(\eta_4 - 1) - \eta_3^2} c(\tau_0) F_{n3}^* e_{kk}' - \frac{\eta_3^2}{(\eta_4 - 1) - \eta_3^2} F_{n3}^* (\frac{1}{n} l_n' e^{\tau_0 M_n} X_n),$$

where e_{kj} is the jth unit vector in R^k. From the proof, the best moments in Proposition 8 are equivalent to the use of the following moments $\frac{1}{n}\big(P_{\alpha n}^* \epsilon_n(\gamma), P_{\tau n}^* \epsilon_n(\gamma), P_{\beta n1}^* \epsilon_n(\gamma) \ldots, P_{\beta n k^*}^* \epsilon_n(\gamma), F_{\alpha n}^*, F_{\beta n}^*\big)' \epsilon_n(\gamma)$. The above vector relates the moments to the skewness and kurtosis.

In the case of normal disturbances, as $\eta_3 = \eta_4 - 3 = 0$, the best moments can be simplified and are equivalent to $\frac{1}{n}\big(\mathbb{W}_n \epsilon_n(\gamma), M_n \epsilon_n(\gamma), P_{\beta n1}^* \epsilon_n(\gamma) \ldots, P_{\beta n k^*}^* \epsilon_n(\gamma), e^{\tau_0 M_n} W_n X_n \beta_0, e^{\tau_0 M_n} X_n\big)' \epsilon_n(\gamma)$. Furthermore, the

[21] If M_n is row-normalized and X_n contains an intercept term, $e^{\tau_0 M_n} l_n = \sum_{j=0}^{\infty} \frac{1}{j!} \tau_0^j M_n^j l_n = e^{\tau_0} l_n$. In this case, $c(\tau_0) = e^{\tau_0}$. Otherwise $e^{\tau_0 M_n} X_n$ generally does not contain an intercept term.

moments $\big(P^*_{\beta n1}\epsilon_n(\gamma)\ldots, P^*_{\beta nk^*}\epsilon_n(\gamma)\big)'\epsilon_n(\gamma)$ can be shown to be redundant given

$$g^{\#}_n(\gamma) = \frac{1}{n}\big(\mathbb{W}_n\epsilon_n(\gamma), M_n\epsilon_n(\gamma), e^{\tau_0 M_n}W_nX_n\beta_0, e^{\tau_0 M_n}X_n\big)'\epsilon_n(\gamma), \tag{16}$$

by an argument similar to the proof of Proposition 8. This result can also be shown by using the generalized Cauchy-Schwarz inequality, as in subsequent section.

The G_n in Proposition 7 can be written as $G_n = \frac{1}{n\sigma_0^2}G'_{2n}G_{1n}$, where

$$G_{1n} = \begin{pmatrix} \frac{\sqrt{2}\sigma_0^2}{2}\,\mathrm{vec}(\mathbb{W}^s_n) & \frac{\sqrt{2}\sigma_0^2}{2}\,\mathrm{vec}(M^s_n) & 0 \\ \sigma_0\mathbb{W}_n e^{\tau_0 M_n}X_n\beta_0 & 0 & -\sigma_0 e^{\tau_0 M_n}X_n \end{pmatrix} \text{ and } G_{2n} = \begin{pmatrix} \frac{\sqrt{2}\sigma_0^2}{2}\omega_n & 0 \\ 0 & \sigma_0 F_n \end{pmatrix}.$$

When ϵ_{ni}'s are normal, $\mu_3 = \mu_4 - 3\sigma_0^4 = 0$. Furthermore, even under non-normal disturbances, if $P_{n1}, ..., P_{n,k_p}$ are chosen to have zero diagonal, then $\omega_{nd} = 0$. For those cases, V_n in Proposition 7 reduces to $V_n = \frac{1}{n}G'_{2n}G_{2n}$. Thus for those cases, $G'_nV_n^{-1}G_n \leq \Lambda_n$ by the generalized Cauchy-Schwarz inequality, where $\Lambda_n = \frac{1}{n\sigma_0^4}G'_{1n}G_{1n}$. As $\mathbb{W}_n$ and M_n both have zero traces, when the moment vector is $g^{\#}_n(\gamma)$ in (16), $G'_nV_n^{-1}G_n = \Lambda_n$. Thus the best moment vector is $g^{\#}_n(\gamma)$ in (16) when ϵ_{ni}'s are normal. When W_n and M_n can commute, the best moment vector, with the restriction that P_{ni}'s have zero diagonals, is[22]

$$g^{\#}_{n,d}(\gamma) = \frac{1}{n}\big(W_n\epsilon_n(\gamma), M_n\epsilon_n(\gamma), e^{\tau_0 M_n}W_nX_n\beta_0, e^{\tau_0 M_n}X_n\big)'\epsilon_n(\gamma). \tag{17}$$

Since $G_{1n} = \frac{\sqrt{2}\sigma_0}{2}C_{1n}$, where C_{1n} is given in (11), the asymptotic VC matrix Λ_n^{-1} for the best GMME in the case of normal disturbances is the same as that for the MLE of γ. It is of interest to note that for the case with non-normal disturbances, when W_n and M_n can commute the QMLE of γ happens to be asymptotically efficient within the class of GMMEs with linear and quadratic moments where the quadratic matrices P_{ni}'s have zero diagonals.

Corollary 1. *Suppose that Assumptions 1, 2, 5 and 12–15 hold.*

(i) When the disturbances ϵ_{ni}'s are normal, for the class of GMMEs with linear and quadratic moments where the quadratic matrices P_{ni}'s have zero traces, the best GMME is the optimal GMME with the moment vector $g^{\#}_n(\gamma)$ in (16);

(ii) When W_n and M_n can commute, for the class of GMMEs with linear and quadratic moments where the quadratic matrices P_{ni}'s have zero diagonals, the best GMME is the optimal GMME with the moment vector $g^{\#}_{n,d}(\gamma)$ in (17).

The best moments in the case of normal disturbances are of interest to be compared with those for the SARAR model. For the latter model, the best instruments are $R_n[X_n, W_nS_n^{-1}X_n\beta_0]$ and the matrices for

[22] When $W_n = M_n$, the moment $\frac{1}{n}\epsilon'_n(\gamma)M_n\epsilon_n(\gamma)$ should not be considered.

the best quadratic moments are $R_nW_nS_n^{-1}R_n^{-1} - I_n\,\mathrm{tr}(W_nS_n^{-1})/n$ and $M_nR_n^{-1} - I_n\,\mathrm{tr}(M_nR_n^{-1})/n$, where $R_n = I_n - \rho_0 M_n$ and $S_n = I_n - \lambda_0 W_n$. Thus, in addition to X_n and W_nX_n, higher order spatially lagged X_n, i.e., $W_n^2X_n$, $W_n^3X_n$, etc., will provide additional information. For the quadratic moments, spatial weights matrices of higher order, namely, W_n^2, W_n^3, etc., from which the average of their diagonal elements is subtracted from each diagonal element, can be used as additional orthogonal conditions. On the other hand, the best instruments and quadratic moments for the MESS(1,1) rely simply on spatial weights matrices W_n and M_n. Note also that when there is no MESS process in the disturbances, the moment vector for the best GMME in the case of normal disturbances can be simply taken as $\frac{1}{n}[\epsilon_n'(\gamma)W_n\epsilon_n(\gamma), \epsilon_n'(\gamma)(W_nX_n, X_n)_{IN}]'$, where $(W_nX_n, X_n)_{IN}$ denotes the independent columns of (W_nX_n, X_n).[23] Thus it has a simple form which does not involve any unknown parameter. By contrast, the moment vector for the best optimal GMME of the SAR model can be taken as $\frac{1}{n}[\epsilon_n'(\gamma)(W_nS_n^{-1} - I_n\,\mathrm{tr}(W_nS_n^{-1})/n)\epsilon_n(\gamma), \epsilon_n'(\gamma)(W_nS_n^{-1}X_n, X_n)_{IN}]'$, which involves the unknown parameter λ_0 in the matrix inverse S_n^{-1}.

There exists a link between the MLE (or QMLE) and moment conditions. The first order conditions for the MLE using the function $Q_n(\gamma)$ can be written as

$$\frac{\partial Q_n(\gamma)}{\partial \alpha} = 2(e^{\tau M_n}W_nX_n\beta)'\epsilon_n(\gamma) + 2\epsilon_n'(\gamma)e^{\tau M_n}W_ne^{-\tau M_n}\epsilon_n(\gamma), \tag{18}$$

$$\frac{\partial Q_n(\gamma)}{\partial \tau} = 2\epsilon_n'(\gamma)M_n\epsilon_n(\gamma) \;\text{ and }\; \frac{\partial Q_n(\gamma)}{\partial \beta} = -2(e^{\tau M_n}X_n)'\epsilon_n(\gamma). \tag{19}$$

Thus the underlying moments integrated by the MLE are also the linear moments with instruments from $e^{\tau_0 M_n}X_n$ and $e^{\tau_0 M_n}W_nX_n$, and the quadratic moments with the matrices $\mathbb{W}_n$ and M_n. The matrix $e^{\tau_0 M_n}$ in front of X_n and W_nX_n is a transformation for the MESS disturbances. When the likelihood function is correctly specified under the normal disturbances, the combinations of linear and quadratic moments in (18)–(19) are the efficient ones. But they might not be so when the likelihood function is only a quasi one. The optimal GMME employs an optimal weighting matrix when using the moments $g_n^{\#}(\gamma)$, but the QMLE might not. Thus a best GMME within the class of linear and quadratic moments can be more efficient asymptotically than the QMLE when the disturbances are non-normal or W_n and M_n cannot commute. This can be shown analytically. Let

$$h_n(\gamma) = \frac{2}{n}\begin{pmatrix} (e^{\tau_0 M_n}W_nX_n\beta_0)'\epsilon_n(\gamma) + \epsilon_n'(\gamma)W_n\epsilon_n(\gamma) \\ \epsilon_n'(\gamma)M_n\epsilon_n(\gamma) \\ -(e^{\tau_0 M_n}X_n)'\epsilon_n(\gamma) \end{pmatrix} = Ag_n^{\#}(\gamma), \text{ with } A = 2\begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -I_k \end{pmatrix}.$$

[23] If W_n is row normalized and X_n contains an intercept, as $W_nl_n = l_n$, only one of the two intercepts should be included in (W_nX_n, X_n).

The $h_n(\gamma)$ and $\frac{1}{n}\frac{\partial Q_n(\gamma)}{\partial \gamma}$ have a similar structure: replacing τ_0 in the components $e^{\tau_0 M_n} W_n X_n \beta_0$ and $e^{\tau_0 M_n} X_n$ in $h_n(\gamma)$ by τ yields $\frac{1}{n}\frac{\partial Q_n(\gamma)}{\partial \gamma}$. It is obvious that $\mathrm{E}\,\frac{\partial h_n(\gamma_0)}{\partial \gamma'} = \frac{1}{n}\mathrm{E}\big(\frac{\partial^2 Q_n(\gamma_0)}{\partial \gamma \partial \gamma'}\big)$ and $\mathrm{E}\big(n h_n(\gamma_0) h_n'(\gamma_0)\big) = \frac{1}{n}\mathrm{E}\big(\frac{\partial Q_n(\gamma_0)}{\partial \gamma}\frac{\partial Q_n(\gamma_0)}{\partial \gamma'}\big)$. Thus, by Proposition 4, the asymptotic VC matrix for the QMLE of γ is equal to $\lim_{n\to\infty}\big(\mathrm{E}\,\frac{\partial h_n(\gamma_0)}{\partial \gamma'}\big)^{-1}\mathrm{E}\big(n h_n(\gamma_0) h_n'(\gamma_0)\big)\big(\mathrm{E}\,\frac{\partial h_n(\gamma_0)}{\partial \gamma}\big)^{-1}$. Therefore,

$$\begin{aligned}\big(\mathrm{E}\,\frac{\partial h_n(\gamma_0)}{\partial \gamma'}\big)^{-1}\mathrm{E}[h_n(\gamma_0)h_n'(\gamma_0)]\big(\mathrm{E}\,\frac{\partial h_n'(\gamma_0)}{\partial \gamma}\big)^{-1} &= \Big[G_n^{\#'}(\gamma_0)A'\big[A\,\mathrm{E}\big(g_n^{\#}(\gamma_0)g_n^{\#'}(\gamma_0)\big)A'\big]^{-1}AG_n^{\#}(\gamma_0)\Big]^{-1} \\ &\geq \Big[G_n^{\#'}(\gamma_0)\big[\mathrm{E}\big(g_n^{\#}(\gamma_0)g_n^{\#'}(\gamma_0)\big)\big]^{-1}G_n^{\#}(\gamma_0)\Big]^{-1},\end{aligned} \tag{20}$$

by the generalized Cauchy-Schwarz inequality, where $G_n^{\#}(\gamma_0) = \mathrm{E}\,\frac{\partial g_n^{\#}(\gamma_0)}{\partial \gamma'}$. The last term above is the asymptotic VC matrix of the feasible optimal GMME with the moment vector $g_n^{\#}(\gamma)$. The inequality in (20) becomes an equality if there is a matrix Δ_{gh} such that $G_n^{\#}(\gamma_0) = \mathrm{E}\big(g_n^{\#}(\gamma_0)g_n^{\#'}(\gamma_0)\big)A'\Delta_{gh}$. From Proposition 6, we have

$$G_n^{\#}(\gamma_0) = \frac{1}{n}\begin{pmatrix} \sigma_0^2\,\mathrm{tr}(\mathbb{W}_n^s\mathbb{W}_n) & \sigma_0^2\,\mathrm{tr}(\mathbb{W}_n^s M_n) & 0 \\ \sigma_0^2\,\mathrm{tr}(M_n^s\mathbb{W}_n) & \sigma_0^2\,\mathrm{tr}(M_n^s M_n) & 0 \\ (e^{\tau_0 M_n}W_nX_n\beta_0, e^{\tau_0 M_n}X_n)'e^{\tau_0 M_n}W_nX_n\beta_0 & 0 & -(e^{\tau_0 M_n}W_nX_n\beta_0, e^{\tau_0 M_n}X_n)'e^{\tau_0 M_n}X_n \end{pmatrix}$$

and

$$\mathrm{E}\big(g_n^{\#}(\gamma_0)g_n^{\#'}(\gamma_0)\big)A' = \frac{2\sigma_0^2}{n}G_n^{\#} + \frac{2}{n^2}\begin{pmatrix} (\mu_4 - 3\sigma_0^4)\,\mathrm{vec_D}'(\mathbb{W}_n)\,\mathrm{vec_D}(\mathbb{W}_n) + \mu_3\,\mathrm{vec_D}'(\mathbb{W}_n)e^{\tau_0 M_n}W_nX_n\beta_0 & 0 & -\mu_3\,\mathrm{vec_D}'(\mathbb{W}_n)e^{\tau_0 M_n}X_n \\ 0 & 0 & 0 \\ \mu_3(e^{\tau_0 M_n}W_nX_n\beta_0, e^{\tau_0 M_n}X_n)'\,\mathrm{vec_D}(\mathbb{W}_n) & 0 & 0 \end{pmatrix}.$$

When $\tau_0 = 0$; W_n and M_n can commute; or $\mu_3 = \mu_4 - 3\sigma_0^4 = 0$, we have $\Delta_{gh} = \frac{n}{2\sigma_0^2}I_{k+2}$. Except for those cases, Δ_{gh} may not exist. As $g_n^{\#}(\gamma)$ in (16) is only a special case of linear and quadratic moments, the best GMME in Proposition 8 can be more efficient asymptotically than the QMLE.

The best moment vector $g_n^*(\gamma)$ and the optimal weighting matrix V_n^{*-1} involve unknown parameters. In practice, $g_n^*(\gamma)$ and V_n^{*-1} can be estimated using initial consistent estimates and a feasible best GMME can be derived. Such a feasible best GMME has the same asymptotic distribution as the best GMME in Proposition 8.

Proposition 9. *Suppose that Assumptions 1, 2, 5 and 12–15 hold. Let $\hat\gamma_n$, $\hat\sigma_n^2$, $\hat\mu_{3n}$ and $\hat\mu_{4n}$ be, respectively, $\sqrt{n}$-consistent estimators of γ_0, σ_0^2, μ_3 and μ_4. The $\hat P_{n1}^*$, …, $\hat P_{n,k^*+4}^*$, $\hat F_n^*$, $\hat F_{1n}^*$, …, $\hat F_{n4}^*$ and $\hat V_n^*$ denote the matrices derived when the unknown parameters in P_{n1}^*, …, P_{n,k^*+4}^*, F_n^*, F_{1n}^*, …, F_{n4}^* and V_n^* are replaced by the corresponding consistent estimators. Then the feasible best GMME $\hat\gamma_{n,f}^* = \min_\gamma \hat g_n^{*'}(\gamma)\hat V_n^{*-1}\hat g_n^*(\gamma)$, where $\hat g_n^*(\gamma) = \frac{1}{n}(\hat P_{n1}^*\epsilon(\gamma), \ldots, \hat P_{n,k^*+4}^*\epsilon(\gamma), \hat F_n^*)'\epsilon(\gamma)$, has the same asymptotic distri-*

bution as $\hat{\gamma}_n^* = \min_\gamma g_n^{*'}(\gamma) V_n^{*-1} g_n^*(\gamma)$.

3.2.2. GMME: Heteroskedastic Case

When the disturbances are independent but may have different variances, the GMME can be consistent when the matrices P_{ni}'s have zero diagonals.[24]

Assumption 16. *The n-dimensional square matrices $P_{n1}, \ldots, P_{n,k_p}$ have zero diagonals and are bounded in both row and column sum norms. Elements of F_n are uniformly bounded constants.*

By taking into account variances of disturbances, the identification condition is similarly derived as that in the homoskedastic case.

Assumption 17. *Suppose that F_n may be written as $F_n = (F_{1n}, F_{2n})$ such that $\lim_{n\to\infty} \frac{1}{n} F_{2n}' e^{\tau M_n} X_n$ is nonsingular for any $\tau \in [-\delta, \delta]$. Furthermore, either 1) $\lim_{n\to\infty} \frac{1}{n} F_{1n}' H_{1n}(\tau) e^{\tau M_n} e^{\eta W_n} X_n \beta_0 \neq 0$ for any $\eta \neq 0$ and for all $\tau \in [-\delta, \delta]$; and, for any $\tau \neq \tau_0$, $\lim_{n\to\infty} \frac{1}{n} \mathrm{tr}(e^{(\tau-\tau_0)M_n'} P_{ni} e^{(\tau-\tau_0)M_n} \Sigma_n) \neq 0$, for some $i \in \{1, \cdots, k_p\}$; or, 2) for any $(\alpha, \tau) \neq (\alpha_0, \tau_0)$, $\lim_{n\to\infty} \frac{1}{n} \beta_0' X_n' e^{(\alpha-\alpha_0)W_n'} e^{\tau M_n'} H_{1n}'(\tau) P_{ni} H_{1n}(\tau) e^{\tau M_n} e^{(\alpha-\alpha_0)W_n} X_n \beta_0 + \frac{1}{n} \mathrm{tr}(e^{-\tau_0 M_n'} e^{(\alpha-\alpha_0)W_n'} e^{\tau M_n'} P_{ni} e^{\tau M_n} e^{(\alpha-\alpha_0)W_n} e^{-\tau_0 M_n} \Sigma_n) \neq 0$, for some $i \in \{1, \cdots, k_p\}$.*

Proposition 10. *Under Assumptions 1, 2, 8, 14, 16 and 17, the GMM estimator $\hat{\gamma}_n$ from the minimization of $g_n'(\gamma) a_n a_n' g_n(\gamma)$ is a consistent estimator of γ_0, and*

$$\sqrt{n}(\hat{\gamma}_n - \gamma_0) \xrightarrow{d} N\big(0, \lim_{n\to\infty} (G_n' a_n a_n' G_n)^{-1} G_n' a_n a_n' V_n a_n a_n' G_n (G_n' a_n a_n' G_n)^{-1}\big),$$

where $V_n = n\,\mathrm{E}[g_n(\gamma_0) g_n'(\gamma_0)] = \frac{1}{n} \begin{pmatrix} \frac{1}{2}\omega_n'\omega_n & 0 \\ 0 & F_n' \Sigma_n F_n \end{pmatrix}$, *and*

$$G_n = \mathrm{E}\,\frac{\partial g_n(\gamma_0)}{\partial \gamma'} = \frac{1}{n} \begin{pmatrix} \frac{1}{2}\omega_n' \,\mathrm{vec}(\Sigma_n^{1/2}(\Sigma_n^{-1}\mathbb{W}_n)^s \Sigma_n^{1/2}) & \frac{1}{2}\omega_n' \,\mathrm{vec}(\Sigma_n^{1/2}(\Sigma_n^{-1}M_n)^s \Sigma_n^{1/2}) & 0 \\ F_n' \mathbb{W}_n e^{\tau_0 M_n} X_n \beta_0 & 0 & -F_n' e^{\tau_0 M_n} X_n \end{pmatrix}$$

with $\omega_n = \big(\mathrm{vec}(\Sigma_n^{1/2} P_{n1}^s \Sigma_n^{1/2}), \ldots, \mathrm{vec}(\Sigma_n^{1/2} P_{n,k_p}^s \Sigma_n^{1/2})\big)$, *under the condition that* $\lim_{n\to\infty} a_n' G_n$ *exists and has the full rank* $k+2$.

The V_n does not involve the third and fourth moments of the disturbances, as the matrices in the quadratic forms of disturbances in $g_n(\gamma_0)$ have zero diagonals. An optimal GMME can also be formulated.

Assumption 18. *The limits of $\frac{1}{n}\omega_n'\omega_n$ and $\frac{1}{n} F_n' \Sigma_n F_n$ exist and are nonsingular.*

A consistent estimator for V_n is the matrix $\hat{V}_n$ derived by replacing the Σ_n in V_n by $\hat{\Sigma}_n = \mathrm{Diag}(\hat{\epsilon}_{n1}^2, \ldots, \hat{\epsilon}_{nn}^2)$, where $\hat{\epsilon}_{ni}$'s are the residuals from an initial GMM estimation. Under Assumption 18, the limiting inverse of V_n exists. Then the objective function for the feasible optimal GMME is $g_n'(\gamma) \hat{V}_n^{-1} g_n(\gamma)$.

[24] $P_{ni} = W_n$; $P_{ni} = M_n$ or $P_{ni} = \big(W_n^2 - \mathrm{Diag}(W_n^2)\big)$ constitute three examples of matrices that could be used.

Proposition 11. *Under Assumptions 1, 2, 8, 14 and 16–18, the feasible optimal GMME $\hat{\gamma}_{n,o}$ from the minimization of $g_n'(\gamma)\hat{V}_n^{-1}g_n(\gamma)$ is a consistent estimator of γ_0, and $\sqrt{n}(\hat{\gamma}_{n,o}-\gamma_0) \xrightarrow{d} N\big(0, \lim_{n\to\infty}(G_n'V_n^{-1}G_n)^{-1}\big)$.*

Note that $\mathrm{tr}\big(\Sigma_n P_{ni}^s \Sigma_n (\Sigma_n^{-1}\mathbb{W}_n)^s\big) = \mathrm{tr}\big(\Sigma_n P_{ni}^s \Sigma_n \big(\Sigma_n^{-1}(\mathbb{W}_n - \mathrm{Diag}(\mathbb{W}_n))\big)^s\big)$ as P_{ni}^s has a zero diagonal and Σ_n is a diagonal matrix, then G_n may be written as $G_n = \frac{1}{n}\begin{pmatrix} \frac{\sqrt{2}}{2}\omega_n & 0 \\ 0 & \Sigma_n^{1/2}F_n \end{pmatrix}' G_{1n}$, where

$$G_{1n} = \begin{pmatrix} \frac{\sqrt{2}}{2}\,\mathrm{vec}\big(\Sigma_n^{1/2}\big(\Sigma_n^{-1}(\mathbb{W}_n - \mathrm{Diag}(\mathbb{W}_n))\big)^s\Sigma_n^{1/2}\big) & \frac{\sqrt{2}}{2}\,\mathrm{vec}(\Sigma_n^{1/2}(\Sigma_n^{-1}M_n)^s\Sigma_n^{1/2}) & 0 \\ \Sigma_n^{-1/2}\mathbb{W}_n e^{\tau_0 M_n}X_n\beta_0 & 0 & -\Sigma_n^{-1/2}e^{\tau_0 M_n}X_n \end{pmatrix}.$$

Thus $G_n'V_n^{-1}G_n \le \Lambda_n$ by the generalized Cauchy-Schwarz inequality, where $\Lambda_n = \frac{1}{n}G_{1n}'G_{1n}$. When the moment vector $g_n(\gamma)$ is equal to $g_n^*(\gamma) = \frac{1}{n}\big[\epsilon_n'(\gamma)\Sigma_n^{-1}(\mathbb{W}_n - \mathrm{Diag}(\mathbb{W}_n))\epsilon_n(\gamma), \epsilon_n'(\gamma)\Sigma_n^{-1}M_n\epsilon_n(\gamma), \epsilon_n'(\gamma)F_n^*\big]'$ with $F_n^* = \Sigma_n^{-1}[\mathbb{W}_n e^{\tau_0 M_n}X_n\beta_0, e^{\tau_0 M_n}X_n]$, $G_n'V_n^{-1}G_n = \Lambda_n$. Therefore, if the variances σ_{ni}^2's can be consistently estimated, e.g., when we have a parametric model for the variances, then we may have a feasible best optimal GMME, for which the moment vector is obtained from replacing the γ_0 in $g_n^*(\gamma)$ by an initial consistent estimator.[25]

If the elements of Σ_n cannot be consistently estimated, we do not have a feasible best GMME, e.g., for the unknown heteroskedastic case, Σ_n with n parameters cannot be consistently estimated. However, we may use the moment vector

$$\hat{g}_{n,d}^{\#}(\gamma) = \frac{1}{n}[\epsilon_n'(\gamma)(\hat{\mathbb{W}}_n - \mathrm{Diag}(\hat{\mathbb{W}}_n))\epsilon_n(\gamma), \epsilon_n'(\gamma)M_n\epsilon_n(\gamma), \epsilon_n'(\gamma)[\hat{\mathbb{W}}_n e^{\hat{\tau}_n M_n}X_n\hat{\beta}, e^{\hat{\tau}_n M_n}X_n]]', \tag{21}$$

and implement a feasible optimal GMM estimation. A special case of interest is when W_n and M_n can commute. In that case, $\hat{g}_{n,d}^{\#}(\gamma)$ reduces to

$$\hat{g}_{n,d}^{\#}(\gamma) = \frac{1}{n}[\epsilon_n'(\gamma)W_n\epsilon_n(\gamma), \epsilon_n'(\gamma)M_n\epsilon_n(\gamma), \epsilon_n'(\gamma)[W_n e^{\hat{\tau}_n M_n}X_n\hat{\beta}_n, e^{\hat{\tau}_n M_n}X_n]]'. \tag{22}$$

It can be shown, as for the proof of Proposition 9, that the optimal GMME using the moment vector $\hat{g}_{n,d}^{\#}(\gamma)$ has the same asymptotic distribution as that using the moment vector $g_{n,d}^{\#}(\gamma)$ in (17). As shown in (18)–(19), the QMLE also integrates those moments in $g_{n,d}^{\#}(\gamma)$. But because of the optimal weighting, the optimal GMME using the moment vector $g_{n,d}^{\#}(\gamma)$ is at least as efficient as the QMLE and generally more efficient than the QMLE asymptotically, according to (20) and arguments similar to those after (20).

[25] For the SARAR model with heteroskedasticity, we have also found that the best instruments are $\Sigma_n^{-1}R_n[X_n, W_nS_n^{-1}X_n\beta_0]$ and the matrices for the quadratic moments are $\Sigma_n^{-1}[R_nW_nS_n^{-1}R_n^{-1} - \mathrm{Diag}(R_nW_nS_n^{-1}R_n^{-1})]$ and $\Sigma_n^{-1}[M_nR_n^{-1} - \mathrm{Diag}(M_nR_n^{-1})]$. For the SAR model with heteroskedasticity, Lin and Lee (2010) have not discussed the possible best GMME by the generalized Cauchy-Schwarz inequality as above.

3.3. On the Inference of Elements in Impact Matrices

Assessing the statistical significance of the effect of a change in a regressor on the dependent variable is one of the main objectives of applied economists. In spatial regressions, as shown in Section 2, one first has to compute the reduced form of the specification and calculate the matrix of partial derivatives of the dependent variable with respect to the concerned regressor in order to figure out the matrix of impacts. Inference regarding causal effects should then be based on this matrix, which, for regressor X_{nk}, is presented in (5). All the elements of this impact matrix are possibly different from each other and performing inference on them would be of value.

For the SAR model, LeSage and Pace (2009) propose a Bayesian Markov chain Monte Carlo approach to produce inference on the scalar summary of effects, namely the average direct, indirect and total impacts. In this paper, we take the classical approach based on the Delta method to perform inference on those elements of the impact matrix. Statistical significance on differences of impacts can also be assessed. For instance, one could be interested in testing if the effect of the kth regressor for observation i on y_{ni} will be the same as of the lth regressor (with l possibly different from k) for individual j on y_{nj}, with j possibly different from i.

Let $\hat{\gamma}_n$ be a $\sqrt{n}$-consistent estimator of γ, and e_{ni} be the ith column of I_n. The impact of $x_{n,jp}$ (pth regressor for individual j) on y_{ni} is estimated to be $e'_{ni}e^{-\hat{\alpha}_n W_n}e_{nj}\hat{\beta}_{np}$, and the effect of $x_{n,sq}$ on y_{nr} is estimated to be $e'_{nr}e^{-\hat{\alpha}_n W_n}e_{ns}\hat{\beta}_{nq}$. Then, by the mean value theorem,

$$\begin{aligned}
&\sqrt{n}[(e'_{ni}e^{-\hat{\alpha}_n W_n}e_{nj}\hat{\beta}_{np} - e'_{nr}e^{-\hat{\alpha}_n W_n}e_{ns}\hat{\beta}_{nq}) - (e'_{ni}e^{-\alpha_0 W_n}e_{nj}\beta_{0p} - e'_{nr}e^{-\alpha_0 W_n}e_{ns}\beta_{0q})] \\
&= A_{1n}\sqrt{n}(\hat{\alpha}_n - \alpha_0, \hat{\beta}_{np} - \beta_{0p}, \hat{\beta}_{nq} - \beta_{0q})' + o_P(1) \\
&\xrightarrow{d} N(0, \lim_{n\to\infty} A_{1n}B_{1n}A'_{1n}),
\end{aligned} \tag{23}$$

where $A_{1n} = [-e'_{ni}e^{-\alpha_0 W_n}W_n e_{nj}\beta_{0p} + e'_{nr}e^{-\alpha_0 W_n}W_n e_{ns}\beta_{0q}, e'_{ni}e^{-\alpha_0 W_n}e_{nj}, -e'_{nr}e^{-\alpha_0 W_n}e_{ns}]$ and B_{1n} is the asymptotic VC matrix of $\sqrt{n}(\hat{\alpha}_n - \alpha_0, \hat{\beta}_{np} - \beta_{0p}, \hat{\beta}_{nq} - \beta_{0q})'$. To test whether the two impacts are equal, we may use the asymptotically standard normal statistic $\sqrt{n}(e'_{ni}e^{-\hat{\alpha}_n W_n}e_{nj}\hat{\beta}_{np} - e'_{nr}e^{-\hat{\alpha}_n W_n}e_{ns}\hat{\beta}_{nq})/(\hat{A}_{1n}\hat{B}_{1n}\hat{A}'_{1n})^{1/2}$ under the null hypothesis, where $\hat{A}_{1n}$ and $\hat{B}_{1n}$ are, respectively, consistent estimates of A_{1n} and B_{1n}. Another example is in testing whether the average direct effect $\frac{1}{n}\operatorname{tr}(e^{\hat{\alpha}_n W_n})\hat{\beta}_{np}$ is significantly different from zero. It can be shown that

$$\begin{aligned}
\frac{1}{\sqrt{n}}\operatorname{tr}(e^{-\hat{\alpha}_n W_n})\hat{\beta}_{np} - \frac{1}{\sqrt{n}}\operatorname{tr}(e^{-\alpha_0 W_n})\beta_{0p} &= A_{2n}\sqrt{n}[\hat{\alpha}_n - \alpha_0, \hat{\beta}_{np} - \beta_{0p}]' + o_P(1) \\
&\xrightarrow{d} N\big(0, \lim_{n\to\infty} A_{2n}B_{2n}A'_{2n}\big),
\end{aligned} \tag{24}$$

where $A_{2n} = [-\frac{1}{n}\operatorname{tr}(e^{-\alpha_0 W_n}W_n)\beta_{0p}, \frac{1}{n}\operatorname{tr}(e^{-\alpha_0 W_n})]$ and B_{2n} is the asymptotic VC matrix of $\sqrt{n}[\hat{\alpha}_n -$

$\alpha_0, \hat{\beta}_{np} - \beta_{0p}]'$. Let $\hat{A}_{2n}$ and $\hat{B}_{2n}$ be, respectively, consistent estimates of A_{2n} and B_{2n}.

Lemma 1. $\sqrt{n}[(e'_{ni}e^{-\hat{\alpha}_n W_n}e_{nj}\hat{\beta}_{np} - e'_{nr}e^{-\hat{\alpha}_n W_n}e_{ns}\hat{\beta}_{nq}) - (e'_{ni}e^{-\alpha_0 W_n}e_{nj}\beta_{0p} - e'_{nr}e^{-\alpha_0 W_n}e_{ns}\beta_{0q})](\hat{A}_{1n}\hat{B}_{1n}\hat{A}'_{1n})^{-1/2} \xrightarrow{d} N(0,1)$ *and* $\frac{1}{\sqrt{n}}[\mathrm{tr}(e^{-\hat{\alpha}_n W_n})\hat{\beta}_{np} - \mathrm{tr}(e^{-\alpha_0 W_n})\beta_{0p}](\hat{A}_{2n}\hat{B}_{2n}\hat{A}'_{2n})^{-1/2} \xrightarrow{d} N(0,1)$.

Several applications of this lemma will be presented in Section 5 which is dedicated to the application of the MESS to figure out the dominant type of outward FDI for Belgium. However, before turning to the empirical application, we first present Monte Carlo experiments which assess the finite sample performance of the MLEs, QMLEs and GMMEs.

4. Monte Carlo Simulations

We consider a MESS(1,1) model with two regressors: $e^{\alpha W_n}y_n = \beta_1 X_{n1} + \beta_2 X_{n2} + u_n, e^{\tau M_n}u_n = \epsilon_n$. The interaction matrix W_n is defined as the 5 nearest neighbors, while we considered two different definitions for M_n. Firstly, $M_n = W_n$ which makes the QMLE consistent even in the presence of unknown heteroskedasticity. Secondly, M_n is defined as a 15 nearest neighbors. In this case, W_n and M_n do not commute and the QMLE will not be consistent in the presence of unknown heteroskedasticity.[26] The elements of X_{n1} and X_{n2} are independently drawn from, respectively, the uniform distribution $U(0,10)$ and the standard normal distribution. For each repetition, the regressors are randomly redrawn.

Three different specifications for the error term are considered. In the first case, the disturbances are i.i.d. normal; in the second case, the disturbances are i.i.d. with a standardized $\Gamma(2,1)$ distribution; in the third case, the disturbances are heteroskedastic, where the heteroskedasticity is defined as the multiplication of a standardized $\Gamma(2,1)$ distribution by the value of the first regressor. In the homoskedastic cases (the first two cases), the variance σ^2 of disturbances is set to keep the signal-to-noise ratio constant. This ratio is defined as the variance of $\beta_1 X_{n1} + \beta_2 X_{n2}$ over the sum of variances of $\beta_1 X_{n1} + \beta_2 X_{n2}$ and ϵ_n (see Liu et al., 2010). In these Monte Carlo experiments, the signal-to-noise ratio is set to 0.5. In the absence of spatial autocorrelation, this ratio would represent a $R^2 = 0.5$. In the heteroskedastic case, the average variances of the disturbances is set to be equal to σ^2.

Two sample sizes, 100 and 254, are considered, corresponding respectively to the number of counties in North Carolina and in Texas. The values of α and τ vary from -2 to 2 by increment of 1 while β_1 and β_2 are set to 1. All the experiments were replicated 1000 times. For these simulations, the GMM estimator is a two step feasible optimal GMME with the first step weighting matrix being an identity matrix. The best

[26] Interaction matrices have been normalized by the spectral radius.

moments are used for the homoskedastic case,[27] while the moment vector for the heteroskedastic case is (21). When there are unknown parameters in the moment conditions, the initial consistent estimator used is the GMME with the moment vector $[\epsilon_n'(\gamma)W_n\epsilon_n, \epsilon_n'(\gamma_n)M_n\epsilon_n(\gamma), \epsilon_n'(\gamma)(W_nX_n, X_n)]'$.

Tables 1–3 summarize the results for the homoskedastic cases with the sample size $n = 100$.[28] We report the bias, standard errors (in italics) and root mean squared error (RMSE) (in bold) for both QML and GMM estimates. When the disturbances are normal, the results for the QML and GMM estimates are very similar in most cases. While the biases for α, β_1 and β_2 are all smaller than 0.1, the bias for τ can be close to 0.3. Relatively large standard error is also observed for τ in some cases. When the disturbances are non-normal, for β_1 and β_2 in all cases and α in most cases, the GMM estimator has smaller standard error than the QML estimator; for τ, the GMM estimator has a smaller standard error only when $W_n = M_n$.

Table 4 reports the estimates of α and τ in the heteroskedastic case. When $W_n = M_n$, the biases of the QML and GMM estimators are very small and their variances have similar magnitudes. When $W_n \neq M_n$, we observe that it is mainly the estimation of τ which is affected by the difference in interaction matrices. Indeed, for both QML and GMM estimation procedures, estimators of α, β_1 and β_2 have similarly small biases and RMSE. For positives values of τ, GMME behaves better than its QMLE counterpart while for negative and null values of τ both estimators behaves similarly. Even though the QMLE is inconsistent theoretically when $W_n \neq M_n$ and $\tau \neq 0$, the largest bias for the QMLE of τ in simulations does not exceed 0.3. For $n = 254$, we do not observe any difference in the behavior of QMLE and GMME.

5. Application to Belgium's outward FDI

To the best of our knowledge, with the recent exceptions of Coughlin and Segev (2000); Blonigen et al. (2007); Baltagi et al. (2007, 2008) and Garretsen and Peeters (2009), the literature on FDI has overlooked the third country effect as a determinant of bilateral FDI. Coughlin and Segev (2000) consider inward FDI for 29 Chinese provinces and find positive and significant spatially autocorrelated error terms (SEM specification). Blonigen et al. (2007) distinguish 4 different types of FDI that multinational enterprises (MNEs) can undertake, summarized in Table 5 (corresponding to Table 1 in Blonigen et al., 2007) and can be identified based on the sign of the spatial lag parameter and of the surrounding-market potential variable.[29]

[27] The moment vector (16) for normal disturbances, and $g_n^*(\gamma)$ in Proposition 8 for non-normal disturbances.

[28] The results for $n = 254$ are similar. We observe that bias, standard errors and thus RMSE are lower when the sample size becomes larger. These results are reported in the supplementary file. Besides, estimates for β_1 and β_2 are only reported for the QML estimator in the homoskedastic case and others are in the supplementary file.

[29] Since the data we have do concern countries and not MNEs, we can only observe the dominant type of MNE behavior in terms of FDI, as the data may contain a mixture of the different motivations for FDI.

Table 1: QML estimation results for $\hat{\alpha}$ and $\hat{\tau}$ for the homoskedastic case with n=100

Results for the $\hat{\alpha}$ estimator

	Results for $W_n = M_n$										Results for $W_n \neq M_n$									
	Normally distributed error terms					Non-normally distributed error terms					Normally distributed error terms					Non-normally distributed error terms				
	α					α					α					α				
	-2	-1	0	1	2	-2	-1	0	1	2	-2	-1	0	1	2	-2	-1	0	1	2
	0.027	0.029	0.027	0.027	0.029	0.013	0.017	0.012	0.011	0.021	0.007	0.007	0.008	0.007	0.008	0.007	0.007	0.008	0.008	0.007
$\tau = -2$	*0.261*	*0.261*	*0.260*	*0.260*	*0.262*	*0.248*	*0.250*	*0.250*	*0.250*	*0.248*	*0.162*	*0.161*	*0.161*	*0.161*	*0.161*	*0.158*	*0.158*	*0.158*	*0.157*	*0.158*
	0.262	**0.262**	**0.261**	**0.262**	**0.263**	**0.248**	**0.250**	**0.251**	**0.251**	**0.249**	**0.162**	**0.161**	**0.161**	**0.161**	**0.161**	**0.158**	**0.159**	**0.158**	**0.158**	**0.158**
	0.002	0.002	0.003	0.002	0.004	0.002	0.000	0.002	0.002	0.003	0.000	0.002	0.003	0.003	0.002	0.005	0.003	0.004	0.005	0.004
$\tau = -1$	*0.143*	*0.143*	*0.142*	*0.143*	*0.143*	*0.136*	*0.136*	*0.136*	*0.137*	*0.136*	*0.128*	*0.127*	*0.127*	*0.127*	*0.127*	*0.127*	*0.125*	*0.126*	*0.128*	*0.126*
	0.143	**0.143**	**0.142**	**0.143**	**0.143**	**0.136**	**0.136**	**0.136**	**0.137**	**0.136**	**0.128**	**0.127**	**0.127**	**0.127**	**0.127**	**0.127**	**0.125**	**0.126**	**0.128**	**0.126**
	0.011	0.012	0.010	0.012	0.010	0.005	0.006	0.007	0.006	0.007	0.007	0.007	0.006	0.006	0.007	0.007	0.008	0.007	0.008	0.007
$\tau = 0$	*0.111*	*0.113*	*0.110*	*0.111*	*0.111*	*0.107*	*0.106*	*0.107*	*0.108*	*0.108*	*0.096*	*0.096*	*0.095*	*0.095*	*0.095*	*0.098*	*0.099*	*0.098*	*0.099*	*0.098*
	0.112	**0.114**	**0.111**	**0.112**	**0.111**	**0.107**	**0.106**	**0.107**	**0.108**	**0.108**	**0.096**	**0.096**	**0.095**	**0.095**	**0.095**	**0.099**	**0.099**	**0.098**	**0.099**	**0.099**
	0.016	0.015	0.016	0.016	0.016	0.007	0.008	0.005	0.008	0.007	0.009	0.009	0.010	0.009	0.010	0.010	0.009	0.008	0.009	0.009
$\tau = 1$	*0.121*	*0.121*	*0.121*	*0.122*	*0.122*	*0.118*	*0.118*	*0.117*	*0.118*	*0.117*	*0.093*	*0.092*	*0.094*	*0.093*	*0.093*	*0.095*	*0.095*	*0.095*	*0.095*	*0.094*
	0.122	**0.122**	**0.123**	**0.123**	**0.123**	**0.118**	**0.118**	**0.117**	**0.118**	**0.118**	**0.093**	**0.092**	**0.094**	**0.093**	**0.093**	**0.095**	**0.095**	**0.095**	**0.095**	**0.094**
	0.014	0.016	0.016	0.014	0.015	0.003	0.006	0.003	0.004	0.005	0.009	0.009	0.009	0.009	0.008	0.008	0.008	0.008	0.008	0.009
$\tau = 2$	*0.118*	*0.118*	*0.118*	*0.117*	*0.118*	*0.116*	*0.115*	*0.116*	*0.116*	*0.116*	*0.091*	*0.091*	*0.091*	*0.091*	*0.090*	*0.091*	*0.091*	*0.091*	*0.091*	*0.091*
	0.119	**0.119**	**0.120**	**0.118**	**0.119**	**0.116**	**0.116**	**0.116**	**0.116**	**0.116**	**0.091**	**0.091**	**0.091**	**0.091**	**0.090**	**0.092**	**0.091**	**0.091**	**0.092**	**0.091**

Results for the $\hat{\tau}$ estimator

	Results for $W_n = M_n$										Results for $W_n \neq M_n$									
	Normally distributed error terms					Non-normally distributed error terms					Normally distributed error terms					Non-normally distributed error terms				
	α					α					α					α				
	-2	-1	0	1	2	-2	-1	0	1	2	-2	-1	0	1	2	-2	-1	0	1	2
	0.016	0.019	0.016	0.016	0.021	0.016	0.020	0.013	0.013	0.023	0.049	0.049	0.050	0.053	0.049	0.075	0.075	0.073	0.075	0.072
$\tau = -2$	*0.325*	*0.323*	*0.324*	*0.325*	*0.326*	*0.314*	*0.313*	*0.315*	*0.315*	*0.313*	*0.382*	*0.378*	*0.384*	*0.380*	*0.381*	*0.381*	*0.381*	*0.380*	*0.382*	*0.379*
	0.326	**0.324**	**0.325**	**0.326**	**0.327**	**0.314**	**0.314**	**0.316**	**0.316**	**0.313**	**0.385**	**0.382**	**0.387**	**0.384**	**0.385**	**0.389**	**0.388**	**0.387**	**0.389**	**0.386**
	0.026	0.026	0.023	0.026	0.025	0.018	0.014	0.017	0.015	0.018	0.108	0.100	0.102	0.104	0.101	0.121	0.116	0.123	0.120	0.122
$\tau = -1$	*0.234*	*0.234*	*0.234*	*0.234*	*0.235*	*0.224*	*0.223*	*0.224*	*0.223*	*0.224*	*0.382*	*0.382*	*0.383*	*0.383*	*0.380*	*0.383*	*0.379*	*0.386*	*0.382*	*0.385*
	0.235	**0.236**	**0.235**	**0.236**	**0.237**	**0.225**	**0.224**	**0.224**	**0.224**	**0.225**	**0.397**	**0.395**	**0.396**	**0.397**	**0.393**	**0.402**	**0.397**	**0.405**	**0.400**	**0.404**
	0.050	0.053	0.051	0.052	0.048	0.035	0.035	0.036	0.033	0.034	0.153	0.152	0.155	0.154	0.156	0.161	0.167	0.162	0.166	0.161
$\tau = 0$	*0.219*	*0.221*	*0.219*	*0.219*	*0.219*	*0.208*	*0.208*	*0.210*	*0.209*	*0.211*	*0.364*	*0.367*	*0.372*	*0.368*	*0.366*	*0.377*	*0.377*	*0.377*	*0.379*	*0.377*
	0.224	**0.227**	**0.225**	**0.225**	**0.224**	**0.211**	**0.211**	**0.213**	**0.212**	**0.214**	**0.395**	**0.397**	**0.403**	**0.399**	**0.398**	**0.410**	**0.412**	**0.410**	**0.414**	**0.409**
	0.075	0.075	0.075	0.075	0.076	0.054	0.055	0.053	0.057	0.054	0.209	0.209	0.209	0.210	0.211	0.214	0.206	0.208	0.208	0.205
$\tau = 1$	*0.229*	*0.229*	*0.229*	*0.230*	*0.229*	*0.218*	*0.218*	*0.219*	*0.217*	*0.217*	*0.398*	*0.397*	*0.397*	*0.397*	*0.399*	*0.395*	*0.394*	*0.393*	*0.394*	*0.395*
	0.241	**0.241**	**0.241**	**0.242**	**0.241**	**0.224**	**0.225**	**0.225**	**0.225**	**0.224**	**0.449**	**0.449**	**0.449**	**0.449**	**0.451**	**0.449**	**0.444**	**0.444**	**0.446**	**0.445**
	0.091	0.092	0.093	0.091	0.092	0.068	0.072	0.069	0.071	0.069	0.283	0.285	0.282	0.280	0.284	0.274	0.281	0.280	0.280	0.281
$\tau = 2$	*0.224*	*0.225*	*0.227*	*0.225*	*0.226*	*0.217*	*0.217*	*0.217*	*0.217*	*0.219*	*0.427*	*0.428*	*0.425*	*0.424*	*0.426*	*0.421*	*0.421*	*0.421*	*0.422*	*0.424*
	0.242	**0.243**	**0.245**	**0.243**	**0.244**	**0.228**	**0.228**	**0.228**	**0.229**	**0.230**	**0.513**	**0.515**	**0.510**	**0.509**	**0.512**	**0.502**	**0.507**	**0.506**	**0.506**	**0.509**

Figures in italics are the standard errors, figures in bold represent RMSE and figures without any layout are the biases.

Table 2: QML estimation results for $\hat{\beta}_1$ and $\hat{\beta}_2$ for the homoskedastic case with n=100

Results for the $\hat{\beta}_1$ estimator

	Results for $W_n = M_n$										Results for $W_n \neq M_n$									
	Normally distributed error terms					Non-normally distributed error terms					Normally distributed error terms					Non-normally distributed error terms				
	α					α					α					α				
	-2	-1	0	1	2	-2	-1	0	1	2	-2	-1	0	1	2	-2	-1	0	1	2
	0.009	0.011	0.009	0.010	0.009	0.002	0.004	0.002	0.002	0.004	0.005	0.006	0.005	0.004	0.005	0.003	0.002	0.004	0.003	0.002
$\tau = -2$	*0.104*	*0.104*	*0.103*	*0.104*	*0.104*	*0.101*	*0.102*	*0.101*	*0.101*	*0.101*	*0.095*	*0.094*	*0.094*	*0.095*	*0.095*	*0.097*	*0.097*	*0.097*	*0.097*	*0.098*
	0.105	**0.104**	**0.104**	**0.104**	**0.105**	**0.101**	**0.102**	**0.101**	**0.101**	**0.101**	**0.095**	**0.094**	**0.094**	**0.095**	**0.095**	**0.097**	**0.097**	**0.097**	**0.097**	**0.098**
	0.006	0.006	0.006	0.006	0.006	0.002	0.002	0.001	0.001	0.000	0.005	0.005	0.005	0.005	0.005	0.000	0.001	0.001	0.001	0.000
$\tau = -1$	*0.084*	*0.083*	*0.083*	*0.083*	*0.083*	*0.084*	*0.084*	*0.083*	*0.084*	*0.083*	*0.093*	*0.092*	*0.092*	*0.093*	*0.092*	*0.095*	*0.094*	*0.094*	*0.094*	*0.094*
	0.084	**0.084**	**0.084**	**0.083**	**0.083**	**0.084**	**0.084**	**0.083**	**0.084**	**0.083**	**0.093**	**0.092**	**0.092**	**0.093**	**0.092**	**0.095**	**0.094**	**0.094**	**0.094**	**0.094**
	0.007	0.008	0.006	0.007	0.007	0.001	0.001	0.001	0.000	0.001	0.005	0.005	0.004	0.005	0.005	0.003	0.003	0.003	0.003	0.002
$\tau = 0$	*0.099*	*0.099*	*0.098*	*0.099*	*0.099*	*0.098*	*0.098*	*0.098*	*0.099*	*0.099*	*0.089*	*0.089*	*0.089*	*0.088*	*0.089*	*0.092*	*0.092*	*0.091*	*0.092*	*0.092*
	0.099	**0.100**	**0.099**	**0.099**	**0.099**	**0.098**	**0.098**	**0.098**	**0.099**	**0.099**	**0.089**	**0.089**	**0.089**	**0.088**	**0.089**	**0.092**	**0.092**	**0.091**	**0.092**	**0.092**
	0.009	0.008	0.009	0.009	0.009	0.001	0.000	0.002	0.000	0.000	0.005	0.005	0.005	0.005	0.005	0.005	0.003	0.003	0.004	0.004
$\tau = 1$	*0.116*	*0.117*	*0.117*	*0.117*	*0.117*	*0.115*	*0.115*	*0.115*	*0.115*	*0.115*	*0.091*	*0.090*	*0.092*	*0.092*	*0.092*	*0.093*	*0.093*	*0.093*	*0.093*	*0.093*
	0.116	**0.117**	**0.117**	**0.118**	**0.117**	**0.115**	**0.115**	**0.115**	**0.115**	**0.115**	**0.092**	**0.091**	**0.092**	**0.092**	**0.092**	**0.094**	**0.094**	**0.093**	**0.093**	**0.093**
	0.007	0.009	0.009	0.007	0.008	0.004	0.000	0.004	0.003	0.002	0.005	0.005	0.004	0.005	0.004	0.003	0.004	0.003	0.004	0.005
$\tau = 2$	*0.116*	*0.116*	*0.116*	*0.115*	*0.116*	*0.115*	*0.114*	*0.114*	*0.115*	*0.114*	*0.090*	*0.090*	*0.091*	*0.090*	*0.090*	*0.090*	*0.090*	*0.090*	*0.090*	*0.090*
	0.116	**0.116**	**0.116**	**0.115**	**0.116**	**0.115**	**0.114**	**0.115**	**0.115**	**0.114**	**0.091**	**0.091**	**0.091**	**0.090**	**0.090**	**0.090**	**0.090**	**0.090**	**0.090**	**0.090**

Results for the $\hat{\beta}_2$ estimator

	Results for $W_n = M_n$										Results for $W_n \neq M_n$									
	Normally distributed error terms					Non-normally distributed error terms					Normally distributed error terms					Non-normally distributed error terms				
	α					α					α					α				
	-2	-1	0	1	2	-2	-1	0	1	2	-2	-1	0	1	2	-2	-1	0	1	2
	0.011	0.012	0.013	0.011	0.010	0.006	0.006	0.007	0.006	0.004	0.010	0.011	0.011	0.005	0.009	0.006	0.008	0.007	0.007	0.009
$\tau = -2$	*0.220*	*0.218*	*0.217*	*0.219*	*0.220*	*0.231*	*0.231*	*0.231*	*0.232*	*0.232*	*0.272*	*0.271*	*0.270*	*0.273*	*0.271*	*0.287*	*0.285*	*0.287*	*0.286*	*0.286*
	0.220	**0.218**	**0.218**	**0.219**	**0.220**	**0.231**	**0.231**	**0.231**	**0.232**	**0.232**	**0.272**	**0.271**	**0.271**	**0.273**	**0.271**	**0.287**	**0.285**	**0.287**	**0.286**	**0.286**
	0.007	0.005	0.009	0.002	0.006	0.008	0.009	0.007	0.009	0.007	0.008	0.010	0.011	0.011	0.009	0.008	0.009	0.004	0.006	0.006
$\tau = -1$	*0.273*	*0.273*	*0.273*	*0.271*	*0.274*	*0.287*	*0.287*	*0.288*	*0.286*	*0.289*	*0.295*	*0.291*	*0.290*	*0.293*	*0.292*	*0.311*	*0.312*	*0.314*	*0.313*	*0.313*
	0.273	**0.273**	**0.274**	**0.271**	**0.274**	**0.287**	**0.287**	**0.288**	**0.287**	**0.289**	**0.295**	**0.291**	**0.290**	**0.294**	**0.292**	**0.311**	**0.312**	**0.314**	**0.313**	**0.313**
	0.009	0.006	0.010	0.007	0.009	0.002	0.001	0.002	0.008	0.000	0.010	0.012	0.014	0.012	0.009	0.005	0.010	0.005	0.003	0.007
$\tau = 0$	*0.304*	*0.305*	*0.300*	*0.304*	*0.303*	*0.328*	*0.329*	*0.329*	*0.326*	*0.327*	*0.305*	*0.307*	*0.307*	*0.302*	*0.306*	*0.331*	*0.331*	*0.331*	*0.331*	*0.331*
	0.304	**0.305**	**0.301**	**0.304**	**0.303**	**0.328**	**0.329**	**0.329**	**0.326**	**0.327**	**0.305**	**0.307**	**0.307**	**0.302**	**0.306**	**0.331**	**0.331**	**0.331**	**0.331**	**0.331**
	0.012	0.012	0.014	0.012	0.013	0.003	0.003	0.005	0.007	0.004	0.013	0.013	0.013	0.013	0.012	0.001	0.003	0.004	0.002	0.001
$\tau = 1$	*0.249*	*0.250*	*0.248*	*0.250*	*0.248*	*0.273*	*0.273*	*0.270*	*0.273*	*0.274*	*0.299*	*0.297*	*0.298*	*0.298*	*0.300*	*0.310*	*0.310*	*0.310*	*0.309*	*0.311*
	0.249	**0.250**	**0.249**	**0.250**	**0.248**	**0.273**	**0.273**	**0.270**	**0.273**	**0.274**	**0.299**	**0.297**	**0.298**	**0.298**	**0.300**	**0.310**	**0.310**	**0.310**	**0.309**	**0.311**
	0.011	0.012	0.012	0.011	0.011	0.001	0.000	0.000	0.001	0.003	0.016	0.014	0.015	0.015	0.014	0.000	0.000	0.002	0.002	0.003
$\tau = 2$	*0.163*	*0.164*	*0.163*	*0.164*	*0.162*	*0.174*	*0.172*	*0.173*	*0.173*	*0.172*	*0.230*	*0.228*	*0.228*	*0.227*	*0.228*	*0.237*	*0.236*	*0.237*	*0.235*	*0.234*
	0.163	**0.165**	**0.163**	**0.165**	**0.162**	**0.174**	**0.172**	**0.173**	**0.173**	**0.172**	**0.230**	**0.228**	**0.228**	**0.227**	**0.229**	**0.237**	**0.236**	**0.237**	**0.235**	**0.234**

Figures in italics are the standard errors, figures in bold represent RMSE and figures without any layout are the biases.

Table 3: GMM estimation results for $\hat{\alpha}$ and $\hat{\tau}$ for the homoskedastic case with n=100

Results for the $\hat{\alpha}$ estimator

	Results for $W_n = M_n$										Results for $W_n \neq M_n$									
	Normally distributed error terms					Non-normally distributed error terms					Normally distributed error terms					Non-normally distributed error terms				
	α					α					α					α				
	-2	-1	0	1	2	-2	-1	0	1	2	-2	-1	0	1	2	-2	-1	0	1	2
	0.010	0.014	0.015	0.014	0.016	0.031	0.032	0.043	0.043	0.042	0.004	0.020	0.022	0.022	0.022	0.016	0.014	0.016	0.014	0.015
$\tau = -2$	*0.240*	*0.246*	*0.244*	*0.244*	*0.244*	*0.213*	*0.214*	*0.415*	*0.421*	*0.414*	*0.160*	*0.173*	*0.175*	*0.175*	*0.174*	*0.152*	*0.154*	*0.156*	*0.153*	*0.154*
	0.240	**0.246**	**0.244**	**0.245**	**0.245**	**0.215**	**0.216**	**0.417**	**0.423**	**0.416**	**0.160**	**0.174**	**0.176**	**0.176**	**0.176**	**0.153**	**0.155**	**0.157**	**0.154**	**0.154**
	0.001	0.001	0.002	0.002	0.002	0.021	0.021	0.022	0.021	0.021	0.001	0.015	0.016	0.014	0.015	0.013	0.011	0.011	0.011	0.011
$\tau = -1$	*0.143*	*0.144*	*0.143*	*0.144*	*0.143*	*0.127*	*0.127*	*0.128*	*0.127*	*0.127*	*0.132*	*0.130*	*0.130*	*0.131*	*0.131*	*0.125*	*0.126*	*0.126*	*0.126*	*0.127*
	0.143	**0.144**	**0.143**	**0.144**	**0.143**	**0.129**	**0.129**	**0.130**	**0.129**	**0.129**	**0.132**	**0.131**	**0.131**	**0.132**	**0.132**	**0.126**	**0.127**	**0.127**	**0.127**	**0.127**
	0.009	0.009	0.008	0.008	0.008	0.012	0.013	0.013	0.013	0.013	0.002	0.010	0.011	0.010	0.011	0.010	0.011	0.011	0.011	0.011
$\tau = 0$	*0.111*	*0.110*	*0.110*	*0.111*	*0.111*	*0.087*	*0.088*	*0.087*	*0.088*	*0.088*	*0.105*	*0.100*	*0.100*	*0.099*	*0.099*	*0.088*	*0.088*	*0.087*	*0.088*	*0.088*
	0.112	**0.111**	**0.110**	**0.111**	**0.111**	**0.088**	**0.089**	**0.088**	**0.089**	**0.089**	**0.105**	**0.100**	**0.100**	**0.100**	**0.100**	**0.088**	**0.089**	**0.088**	**0.089**	**0.089**
	0.001	0.008	0.009	0.009	0.008	0.013	0.008	0.007	0.008	0.008	0.004	0.007	0.008	0.008	0.008	0.008	0.010	0.009	0.009	0.009
$\tau = 1$	*0.121*	*0.120*	*0.120*	*0.120*	*0.119*	*0.098*	*0.090*	*0.090*	*0.090*	*0.090*	*0.100*	*0.098*	*0.096*	*0.096*	*0.096*	*0.080*	*0.079*	*0.079*	*0.079*	*0.079*
	0.121	**0.120**	**0.121**	**0.121**	**0.120**	**0.099**	**0.090**	**0.090**	**0.090**	**0.090**	**0.100**	**0.098**	**0.097**	**0.096**	**0.096**	**0.081**	**0.080**	**0.079**	**0.079**	**0.079**
	0.003	0.004	0.005	0.006	0.004	0.004	0.004	0.004	0.004	0.004	0.003	0.007	0.006	0.006	0.006	0.009	0.007	0.006	0.007	0.007
$\tau = 2$	*0.117*	*0.117*	*0.118*	*0.118*	*0.117*	*0.089*	*0.089*	*0.089*	*0.089*	*0.088*	*0.097*	*0.094*	*0.094*	*0.094*	*0.094*	*0.077*	*0.079*	*0.078*	*0.078*	*0.077*
	0.117	**0.117**	**0.118**	**0.118**	**0.117**	**0.089**	**0.089**	**0.089**	**0.089**	**0.088**	**0.097**	**0.095**	**0.094**	**0.094**	**0.094**	**0.077**	**0.079**	**0.078**	**0.078**	**0.077**

Results for the $\hat{\tau}$ estimator

	Results for $W_n = M_n$										Results for $W_n \neq M_n$									
	Normally distributed error terms					Non-normally distributed error terms					Normally distributed error terms					Non-normally distributed error terms				
	α					α					α					α				
	-2	-1	0	1	2	-2	-1	0	1	2	-2	-1	0	1	2	-2	-1	0	1	2
	0.018	0.024	0.024	0.022	0.026	0.037	0.039	0.052	0.050	0.050	0.051	0.092	0.098	0.099	0.102	0.073	0.072	0.074	0.073	0.072
$\tau = -2$	*0.300*	*0.307*	*0.305*	*0.306*	*0.306*	*0.285*	*0.287*	*0.489*	*0.496*	*0.489*	*0.397*	*0.469*	*0.477*	*0.478*	*0.477*	*0.396*	*0.397*	*0.398*	*0.391*	*0.397*
	0.301	**0.308**	**0.306**	**0.307**	**0.307**	**0.288**	**0.290**	**0.492**	**0.499**	**0.491**	**0.401**	**0.478**	**0.487**	**0.488**	**0.488**	**0.403**	**0.404**	**0.405**	**0.398**	**0.403**
	0.026	0.028	0.029	0.028	0.029	0.042	0.042	0.042	0.041	0.041	0.070	0.127	0.131	0.126	0.131	0.092	0.090	0.087	0.088	0.090
$\tau = -1$	*0.236*	*0.236*	*0.236*	*0.237*	*0.234*	*0.220*	*0.220*	*0.221*	*0.220*	*0.220*	*0.415*	*0.442*	*0.440*	*0.441*	*0.442*	*0.404*	*0.401*	*0.406*	*0.401*	*0.407*
	0.237	**0.238**	**0.238**	**0.238**	**0.236**	**0.224**	**0.224**	**0.225**	**0.223**	**0.224**	**0.421**	**0.460**	**0.459**	**0.459**	**0.461**	**0.415**	**0.411**	**0.415**	**0.411**	**0.417**
	0.048	0.047	0.046	0.047	0.048	0.041	0.044	0.044	0.043	0.044	0.086	0.152	0.153	0.152	0.152	0.115	0.113	0.115	0.113	0.117
$\tau = 0$	*0.222*	*0.222*	*0.220*	*0.222*	*0.220*	*0.203*	*0.204*	*0.204*	*0.204*	*0.203*	*0.442*	*0.420*	*0.421*	*0.421*	*0.420*	*0.407*	*0.404*	*0.398*	*0.404*	*0.401*
	0.227	**0.227**	**0.225**	**0.227**	**0.225**	**0.207**	**0.209**	**0.209**	**0.208**	**0.208**	**0.451**	**0.447**	**0.448**	**0.447**	**0.447**	**0.423**	**0.419**	**0.414**	**0.419**	**0.418**
	0.053	0.065	0.065	0.065	0.065	0.065	0.055	0.053	0.055	0.053	0.093	0.173	0.178	0.183	0.182	0.138	0.135	0.140	0.138	0.136
$\tau = 1$	*0.221*	*0.231*	*0.231*	*0.232*	*0.230*	*0.212*	*0.210*	*0.209*	*0.210*	*0.209*	*0.483*	*0.458*	*0.447*	*0.436*	*0.435*	*0.421*	*0.416*	*0.418*	*0.418*	*0.416*
	0.227	**0.240**	**0.240**	**0.240**	**0.239**	**0.221**	**0.217**	**0.216**	**0.217**	**0.216**	**0.492**	**0.490**	**0.481**	**0.473**	**0.472**	**0.443**	**0.438**	**0.441**	**0.440**	**0.437**
	0.076	0.079	0.081	0.079	0.077	0.064	0.065	0.065	0.063	0.063	0.134	0.193	0.196	0.196	0.195	0.208	0.183	0.187	0.185	0.192
$\tau = 2$	*0.230*	*0.230*	*0.231*	*0.231*	*0.231*	*0.210*	*0.211*	*0.210*	*0.211*	*0.211*	*0.524*	*0.466*	*0.464*	*0.452*	*0.454*	*0.443*	*0.440*	*0.436*	*0.436*	*0.435*
	0.242	**0.244**	**0.245**	**0.244**	**0.243**	**0.220**	**0.221**	**0.220**	**0.220**	**0.220**	**0.541**	**0.504**	**0.503**	**0.492**	**0.494**	**0.490**	**0.477**	**0.474**	**0.473**	**0.475**

Figures in italics are the standard errors, figures in bold represent RMSE and figures without any layout are the biases. We observe that for $W_n = M_n$, non-normal disturbances and $\tau = -2$, the standard errors corresponding to $\alpha = 0$, -1 and 2 are significantly larger than those for $\alpha = -2$ and 1. These are caused by outliers. We have computed a robust measure of the standard deviation, the mean absolute deviation, and find that the values have similar magnitudes for different values of α when $\tau = -2$.

Table 4: Estimation results for $\hat{\alpha}$ and $\hat{\tau}$ for the heteroskedastic case with n=100

Results for the $\hat{\alpha}$ estimator

	Results for QML estimation										Results for GMM estimation									
	Heteroskedastic with $W_n = M_n$					Heteroskedastic with $W_n \neq M_n$					Heteroskedastic with $W_n = M_n$					Heteroskedastic with $W_n \neq M_n$				
	α					α					α					α				
	-2	-1	0	1	2	-2	-1	0	1	2	-2	-1	0	1	2	-2	-1	0	1	2
	0.004	0.006	0.008	0.003	0.004	0.000	0.001	0.000	0.000	0.000	0.012	0.014	0.016	0.016	0.016	0.016	0.008	0.009	0.009	0.003
$\tau = -2$	*0.154*	*0.154*	*0.156*	*0.154*	*0.155*	*0.114*	*0.113*	*0.114*	*0.113*	*0.114*	*0.153*	*0.155*	*0.155*	*0.155*	*0.155*	*0.110*	*0.111*	*0.112*	*0.112*	*0.110*
	0.155	**0.155**	**0.157**	**0.155**	**0.156**	**0.114**	**0.114**	**0.114**	**0.113**	**0.114**	**0.153**	**0.155**	**0.156**	**0.156**	**0.156**	**0.111**	**0.112**	**0.113**	**0.113**	**0.110**
	0.002	0.001	0.000	0.001	0.002	0.001	0.001	0.001	0.000	0.001	0.006	0.007	0.007	0.007	0.006	0.007	0.003	0.003	0.003	0.003
$\tau = -1$	*0.065*	*0.064*	*0.064*	*0.064*	*0.063*	*0.067*	*0.067*	*0.067*	*0.067*	*0.067*	*0.065*	*0.065*	*0.065*	*0.065*	*0.065*	*0.079*	*0.064*	*0.064*	*0.064*	*0.064*
	0.065	**0.064**	**0.064**	**0.064**	**0.063**	**0.067**	**0.067**	**0.067**	**0.067**	**0.067**	**0.066**	**0.066**	**0.066**	**0.066**	**0.065**	**0.079**	**0.064**	**0.064**	**0.064**	**0.064**
	0.000	0.001	0.000	0.001	0.000	0.002	0.002	0.001	0.002	0.002	0.002	0.002	0.002	0.002	0.002	0.000	0.004	0.004	0.005	0.004
$\tau = 0$	*0.047*	*0.047*	*0.047*	*0.046*	*0.047*	*0.046*	*0.047*	*0.047*	*0.047*	*0.046*	*0.049*	*0.049*	*0.049*	*0.049*	*0.049*	*0.060*	*0.049*	*0.047*	*0.047*	*0.047*
	0.047	**0.047**	**0.047**	**0.046**	**0.047**	**0.046**	**0.047**	**0.047**	**0.047**	**0.046**	**0.049**	**0.049**	**0.049**	**0.049**	**0.049**	**0.060**	**0.049**	**0.047**	**0.047**	**0.047**
	0.000	0.000	0.000	0.001	0.001	0.002	0.003	0.002	0.002	0.002	0.004	0.004	0.004	0.005	0.004	0.003	0.006	0.004	0.007	0.006
$\tau = 1$	*0.064*	*0.064*	*0.065*	*0.064*	*0.064*	*0.056*	*0.056*	*0.056*	*0.056*	*0.056*	*0.066*	*0.066*	*0.066*	*0.066*	*0.066*	*0.062*	*0.056*	*0.056*	*0.057*	*0.056*
	0.064	**0.064**	**0.065**	**0.064**	**0.064**	**0.056**	**0.056**	**0.056**	**0.056**	**0.056**	**0.066**	**0.066**	**0.066**	**0.067**	**0.066**	**0.062**	**0.056**	**0.056**	**0.057**	**0.057**
	0.001	0.002	0.001	0.002	0.000	0.003	0.003	0.003	0.003	0.002	0.004	0.000	0.004	0.004	0.004	0.004	0.006	0.005	0.008	0.005
$\tau = 2$	*0.067*	*0.066*	*0.066*	*0.067*	*0.067*	*0.061*	*0.060*	*0.060*	*0.060*	*0.060*	*0.070*	*0.067*	*0.069*	*0.067*	*0.069*	*0.065*	*0.061*	*0.059*	*0.059*	*0.059*
	0.067	**0.066**	**0.066**	**0.067**	**0.067**	**0.061**	**0.060**	**0.060**	**0.060**	**0.060**	**0.070**	**0.067**	**0.069**	**0.067**	**0.069**	**0.065**	**0.061**	**0.060**	**0.060**	**0.060**

Results for the $\hat{\tau}$ estimator

	Results for QML estimation										Results for GMM estimation									
	Heteroskedastic with $W_n = M_n$					Heteroskedastic with $W_n \neq M_n$					Heteroskedastic with $W_n = M_n$					Heteroskedastic with $W_n \neq M_n$				
	α					α					α					α				
	-2	-1	0	1	2	-2	-1	0	1	2	-2	-1	0	1	2	-2	-1	0	1	2
	0.002	0.005	0.008	0.001	0.004	0.079	0.079	0.072	0.077	0.073	0.005	0.002	0.001	0.002	0.001	0.037	0.050	0.055	0.054	0.076
$\tau = -2$	*0.238*	*0.238*	*0.237*	*0.237*	*0.237*	*0.365*	*0.363*	*0.365*	*0.365*	*0.364*	*0.230*	*0.234*	*0.235*	*0.234*	*0.235*	*0.333*	*0.360*	*0.363*	*0.363*	*0.383*
	0.238	**0.238**	**0.237**	**0.237**	**0.237**	**0.373**	**0.372**	**0.372**	**0.373**	**0.371**	**0.231**	**0.234**	**0.235**	**0.234**	**0.235**	**0.335**	**0.363**	**0.367**	**0.367**	**0.391**
	0.012	0.012	0.015	0.014	0.011	0.121	0.124	0.119	0.124	0.126	0.022	0.021	0.021	0.021	0.023	0.055	0.116	0.120	0.119	0.120
$\tau = -1$	*0.186*	*0.186*	*0.186*	*0.186*	*0.187*	*0.359*	*0.360*	*0.360*	*0.359*	*0.357*	*0.187*	*0.190*	*0.190*	*0.190*	*0.189*	*0.339*	*0.350*	*0.351*	*0.351*	*0.351*
	0.187	**0.186**	**0.187**	**0.187**	**0.187**	**0.379**	**0.381**	**0.379**	**0.380**	**0.379**	**0.189**	**0.191**	**0.191**	**0.191**	**0.190**	**0.343**	**0.369**	**0.371**	**0.370**	**0.371**
	0.029	0.027	0.027	0.023	0.024	0.149	0.154	0.152	0.153	0.153	0.039	0.042	0.042	0.042	0.042	0.023	0.139	0.154	0.154	0.154
$\tau = 0$	*0.182*	*0.179*	*0.181*	*0.179*	*0.181*	*0.347*	*0.349*	*0.349*	*0.348*	*0.346*	*0.187*	*0.187*	*0.187*	*0.187*	*0.187*	*0.367*	*0.349*	*0.346*	*0.343*	*0.346*
	0.184	**0.181**	**0.183**	**0.181**	**0.182**	**0.377**	**0.381**	**0.380**	**0.380**	**0.378**	**0.191**	**0.191**	**0.191**	**0.191**	**0.191**	**0.368**	**0.376**	**0.379**	**0.376**	**0.379**
	0.045	0.045	0.044	0.043	0.039	0.193	0.192	0.194	0.193	0.193	0.057	0.060	0.060	0.060	0.060	0.030	0.174	0.176	0.179	0.178
$\tau = 1$	*0.188*	*0.188*	*0.189*	*0.187*	*0.187*	*0.371*	*0.371*	*0.370*	*0.372*	*0.372*	*0.195*	*0.195*	*0.195*	*0.195*	*0.195*	*0.424*	*0.369*	*0.355*	*0.350*	*0.353*
	0.194	**0.193**	**0.194**	**0.192**	**0.191**	**0.418**	**0.417**	**0.418**	**0.419**	**0.419**	**0.203**	**0.204**	**0.204**	**0.204**	**0.204**	**0.425**	**0.408**	**0.396**	**0.393**	**0.395**
	0.062	0.060	0.062	0.061	0.060	0.265	0.262	0.260	0.264	0.263	0.071	0.069	0.075	0.072	0.075	0.091	0.175	0.199	0.210	0.200
$\tau = 2$	*0.189*	*0.186*	*0.190*	*0.190*	*0.190*	*0.406*	*0.406*	*0.404*	*0.404*	*0.402*	*0.198*	*0.194*	*0.197*	*0.184*	*0.197*	*0.453*	*0.416*	*0.377*	*0.376*	*0.379*
	0.199	**0.195**	**0.200**	**0.200**	**0.199**	**0.485**	**0.484**	**0.480**	**0.483**	**0.481**	**0.210**	**0.206**	**0.210**	**0.198**	**0.210**	**0.462**	**0.451**	**0.426**	**0.430**	**0.429**

Figures in italics are the standard errors, figures in bold represent RMSE and figures without any layout are the biases.

Table 5: Expected signs for spatial lag and surrounding-market potential variables

FDI Motivation	Sign of spatial lag	Sign of surrounding-market potential variable
Pure horizontal	0	0
Export-platform	−	+
Pure vertical	−	0
Vertical specialization	+	0

Source: Blonigen et al. (2007).

MNEs can firstly embark in FDI for market access reasons and avoidance of high trade or tariff costs in a host country. This is horizontal FDI. If trade barriers between the parent country (where the MNE is located) and host country (where the MNE would like to make its products available) are too high, the MNE could decide to build a plant in the latter country to avoid export costs but at the expense of building a new production plant. Blonigen et al. (2007) note that no spatial autocorrelation between FDI should be observed since MNEs make independent decisions about serving a market either through exports or affiliate sales. Besides, for this basic form of FDI, we do not expect any market potential effect of host country since the MNE looks for access to the considered market only.

A second motivation for FDI occurs if trade barriers between a set of destination markets are lower than trade frictions between these destination markets and the parent country. In that setup, a MNE could decide to build a plant in a host country, export to other markets and facing lower trade costs only. This type of FDI is called export-platform. As the MNE will not build a production plant in each host country, we expect a negative spatial autocorrelation between neighboring FDI locations. However, we anticipate a positive effect of the surrounding-market potential variable since the MNE will locate its new plant in the host country which has access to the largest surrounding market.

MNEs will make vertical FDI if they want to access to cheaper factor inputs for their products. In its simplest form, namely pure vertical, host countries are in competition in terms of input factor prices to receive FDI. Hence, we expect a negative spatial autocorrelation between FDI. However, since the product is shipped back to the parent country to be further processed, no effect from surrounding-market potential is foreseen. A more complex form of vertical FDI has been developed by Davies (2005) and Baltagi et al. (2007). Within that framework, named vertical specialization, the MNE decides to split its vertical chain of production among possibly several host countries, to benefit from the comparative advantage of the hosts. In such a framework, according to Blonigen et al. (2007), we should observe positive spatial autocorrelation due to possible agglomeration forces such as the presence of immobile resources, since the suppliers' presence in neighboring host countries is likely to increase FDI to a particular market. However, for the same reason

as in pure vertical FDI, we do not predict any surrounding-market effect.

Blonigen et al. (2007) use outbound US FDI to 35 countries over the period 1983 to 1998 to test the dominant type of FDI which characterizes US MNEs. Even though they found a positive and significant spatial dependence effect, the authors acknowledge the fragility of their results with respect to the countries considered. Besides, significance of the surrounding market effect variable is affected by the presence of individual effects in the regression. Garretsen and Peeters (2009) also test the dominant motivation for FDI using outward Dutch FDI to 19 countries from 1984 to 2004. When analyzing their complete sample, they find a positive and significant market potential effect but also positive and significant spatial autocorrelation among FDI.

Our contribution to this literature is threefold. Firstly, we analyze the dominant pattern of Belgium's outward FDI using a modified gravity equation which, in addition to traditional determinants found in the literature, also captures effects of spatial interactions and market potential. We secondly compare results using a MESS(1,1) and a SARAR specification and highlight the similarities in terms of economic interpretations of these two models. We finally apply the lemma concerning inference to assess statistical significance of elements of impact matrices of FDI's determinants.

5.1. Data and empirical specification

This application concerns Belgium's outward FDI into 35 countries in 2009. These 35 host countries belong either to OECD or European Union and represent 94% of Belgium's total outward FDI.[30]

The modified gravity to be estimated is presented in (25).

$$LFDI_i = \beta_1 + \beta_2\, LGDP_i + \beta_3\, LPOP_i + \beta_4\, OECD_i + \beta_5\, LDIS_i + \beta_6\, TARIFFS_i + \beta_7\, MP_i + \epsilon_i. \qquad (25)$$

$LFDI_i$ is the stock of outward FDI (in logs) from Belgium to host country i. FDI stocks were extracted from the OECD International Direct Investment Statistics. The set of regressors includes host GDP in logs ($LGDP$), host population in logs ($LPOP$), an OECD dummy which captures an OECD effect, the bilateral distance between Belgium and country i expressed in logs ($LDIS$) and a measure of trade costs which corresponds to the weighted mean of applied tariffs on all products, as defined by the World bank WDI database and labeled as $TARIFFS$. The last exogenous regressor is the surrounding-market potential variable, MP. We follow a similar approach to Blonigen et al. (2007) in the definition of this variable. For

[30] The countries considered are: Australia, Austria, Bulgaria, Canada, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Japan, South Korea, Latvia, Lithuania, Luxembourg, Mexico, Netherlands, New Zealand, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom and United States of America.

host country i, we define the market potential as the sum of inverse-distance weighted log-GDPs of all other $k \neq i$ countries in the world for which we could obtain GDP data (this amounts to 183 countries). The only difference from Blonigen et al. (2007) is that these authors use the log of the inverse-distance weighted GDP to measure surrounding market potential.[31] $LGDP$, $LPOP$ and $TARIFFS$ all come from the World Bank WDI database while bilateral distances and distances used to construct the MP variable come from CEPII's databases. Finally, all the concerned variables are expressed in constant USD of 2000. Some descriptive statistics of the data are presented in Table 6.

Table 6: Descriptive statistics for the data

	Mean	Std dev	Min	Max
$LFDI$	8.394	1.997	4.410	11.851
$LGDP$	25.973	1.775	22.782	30.048
$LPOP$	16.390	1.450	13.118	19.542
$OECD$	0.857	0.355	0	1
$LDIS$	7.337	1.157	5.154	9.853
$TARIFFS$	1.877	1.369	0.990	8.930
MP	1.364	0.491	0.356	2.257

Accounting for spatial autocorrelation in FDI requires the setup of an interaction scheme, modeled through the interaction (spatial weights) matrix W_n. In this application, we follow Blonigen et al. (2007) and use an inverse arc-distance between capitals to model interactions between host countries. However, we do not multiply the weights by the shortest distance between capitals as done in Blonigen et al. (2007) since we do not row-normalize our weight matrix but instead use the spectral radius to standardize the matrix.[32] This approach is advocated by Baltagi et al. (2008) who argue that row-normalizing a distance based interaction matrix converts absolute distance-based interactions to relative distance-based and thus changes the information content of the interaction scheme.[33] In addition, we control for the presence of residual spatial autocorrelation in the error terms. We consider the same interaction matrix for both MESS processes. As shown in Section 3, the QMLE can be consistent in presence of unknown heteroskedasticity.

Table 7 summarizes the results of different econometric specifications which extend (25). Columns 2–7 present estimation results respectively for homoskedastic SARAR (by QML), homoskedastic MESS(1,1) (by QML), homoskedastic MESS(1,1) (by optimal GMM with the moment vector $\hat{g}^{\#}_{n,d}(\gamma)$ in Proposition 9),

[31] This difference in the position of the logarithm is motivated by the fact that as the host GDP enters equation (25) in logarithms, we believe the surrounding market variable should also be based on logged GDP. Also,Garretsen and Peeters (2009) construct their surrounding-market potential variable in a different way since they only consider the GDP of all host countries in the sample.

[32] As each weight will be multiplied by a common factor, the spectral radius will also be multiplied by this factor, implying that the normalized matrix will be the same, no matter if the interaction matrix is initially rescaled or not.

[33] For further information concerning matrix normalizations, interested readers may consult Kelejian and Prucha (2010).

heteroskedastic SARAR (by optimal GMM[34]), heteroskedastic MESS(1,1) (by QML) and heteroskedastic MESS(1,1) (by optimal GMM with the moment vector $\hat{g}^{\#}_{n,d}(\gamma)$ in (22)).

Let us first note that the quasi maximum likelihood and GMM estimation of the MESS(1,1) with homoskedastic and heteroskedastic disturbances provide similar results for both estimated values and standard errors. The second result we would like to pinpoint relates to the sign of the parameter capturing interactions between observations. We observe a negative λ for both SARAR specifications (homoskedastic and heteroskedastic) while the MESS(1,1) provides a positive value for α. Thus, a negative λ translates in a positive α.[35] Finally, there is no significant spatial autocorrelation left in the error terms.[36]

The computation of matrices of impacts of changes in determinants on FDI is required to be able to give conclusions regarding the dominant type of FDI characterizing Belgium. Indeed, as MESS(1,1) and SARAR are estimated under implicit form (see (2) and (3)), we need to compute their associated reduced form and then calculate the matrix of partial derivatives with respect to each explanatory variable to get impact matrices. For the MESS(1,1), this impact matrix for regressor X_{nk} is shown in (5).

To compare MESS(1,1) and SARAR results, we report in Table 8 the average direct effect and the average total effect for each of the explanatory variables for heteroskedastic SARAR and MESS(1,1), estimated both by QML and GMM. The average direct effect is computed as the average of diagonal elements of the impact matrix while the average total effect is defined as the mean of the row-sum of its elements $\left(\frac{1}{n} l_n' \Xi^{y_n}_{X_{nk}} l_n\right)$. In terms of impacts on the dependent variable, the main focus for economists, we observe a strong similarity of impacts produced by the two specifications even though parameters capturing spatial dependence are completely different from each other.

The lemma derived in Section 3 allows performing inference on elements of the impact matrices of the MESS(1,1). Table 9 summarizes inference results performed on different (functions of) elements of these impact matrices, based on the heteroskedastic MESS(1,1) estimated by GMM. The first row analyzes the significance of average direct effects. The results indicate a non-significant elasticity of surrounding-market potential on FDI. This result, combined with a negative spatial autocorrelation, points to the dominance of pure vertical type of FDI. To the best of our knowledge, this application is the first to indicate such a clear cut result. One possible explanation of this result lies in the production costs faced by Belgian multinationals

[34] In the moment vector, the instruments for the linear moments are $\hat{R}_n[X_n, W_n\hat{S}_n^{-1}X_n\hat{\beta}_n]$ and the matrices for the quadratic moments are $\hat{R}_n W_n \hat{S}_n^{-1}\hat{R}_n^{-1} - \mathrm{Diag}(\hat{R}_n W_n \hat{S}_n^{-1}\hat{R}_n^{-1})$ and $M_n\hat{R}_n^{-1} - \mathrm{Diag}(M_n\hat{R}_n^{-1})$, where $\hat{R}_n = I_n - \hat{\rho}_n M_n$, $\hat{S}_n = I_n - \hat{\lambda}_n W_n$ and $(\hat{\lambda}_n, \hat{\rho}_n, \hat{\beta}_n')'$ is an initial GMME.

[35] This difference comes from the definitions of these two models, as shown in Section 2.

[36] In the supplement file, we report the estimation results of the MESS(1,0). All results are qualitatively the same.

Table 7: Estimation results for different specifications

	(1)	(2)	(3)	(4)	(5)	(6)
Cons.	-2.438	-3.021	-5.367	-1.266	-3.023	-2.819
	(4.582)	(4.660)	(4.564)	(3.899)	(4.288)	(4.274)
$LGDP$	1.106^{***}	1.114^{***}	1.325^{***}	1.089^{***}	1.113^{***}	1.110^{***}
	(0.243)	(0.246)	(0.237)	(0.220)	(0.241)	(0.240)
$LPOP$	-0.584^{**}	-0.585^{**}	-0.803^{***}	-0.591^{**}	-0.584^{**}	-0.586^{**}
	(0.243)	(0.246)	(0.237)	(0.235)	(0.254)	(0.252)
$OECD$	1.064^{*}	1.037^{*}	0.912^{*}	1.023^{*}	1.037^{*}	1.029^{*}
	(0.548)	(0.551)	(0.534)	(0.605)	(0.611)	(0.613)
$LDIS$	-1.245^{***}	-1.199^{***}	-1.177^{***}	-1.293^{***}	-1.199^{***}	-1.206^{***}
	(0.223)	(0.220)	(0.220)	(0.209)	(0.200)	(0.200)
$TARIFFS$	0.107	0.106	0.151	0.108	0.106	0.106
	(0.112)	(0.113)	(0.110)	(0.084)	(0.084)	(0.083)
MP	1.275	1.212	1.534	1.156	1.212	1.183
	(1.079)	(1.105)	(1.095)	(1.147)	(1.186)	(1.185)
Spat auto	-0.331^{**}	0.265^{**}	0.284^{***}	-0.335^{**}	0.265^{**}	0.264^{**}
in y	(0.153)	(0.109)	(0.107)	(0.173)	(0.121)	(0.121)
Spat auto	0.015	-0.004	-0.136	0.282	-0.004	-0.024
in errors	(0.530)	(0.516)	(0.527)	(0.598)	(0.419)	(0.418)
n	35	35	35	35	35	35

Standard errors between brackets; (1) is homoskedastic SARAR, (2) is homo. MESS(1,1) by QML, (3) is homo. MESS(1,1) by GMM, (4) is heteroskedastic SARAR, (5) is hetero. MESS(1,1) by QML and (6) is hetero. MESS(1,1) by GMM; *, ** and *** correspond to significance at the 10%, 5% and 1% respectively.

Table 8: Comparison of average direct effects and average total effects

	Average direct effects			Average total effects		
	SARAR	MESS(1,1) QML	MESS(1,1) GMM	SARAR	MESS(1,1) QML	MESS(1,1) GMM
$LGDP$	1.094	1.116	1.113	0.884	0.921	0.919
$LPOP$	-0.595	-0.586	-0.587	-0.480	-0.484	-0.485
$OECD$	1.029	1.039	1.031	0.831	0.858	0.855
$LDIS$	-1.300	-1.202	-1.209	-1.050	-0.992	-0.998
$TARIFFS$	0.109	0.107	0.106	0.088	0.088	0.088
M_PO	1.163	1.214	1.185	0.939	1.001	0.979

Effects are computed from estimation results of heteroskedastic SARAR and MESS(1,1) (estimated by QML and GMM).

in Belgium. Indeed, labor costs in Belgium are amongst the highest in Europe.[37] Besides, determinants of the traditional gravity equation have the expected sign. We observe a positive and significant elasticity of GDP, which captures the wealth effect, while elasticities of population and bilateral distance are found to be negative. The OECD dummy is found to be significant at the 10% level. Finally, the tariffs variable is found to be non-significant which can be explained by the homogeneity of the sample.

[37] See Eurostat database on labor costs.

The second row presents inference on the indirect effect of Austria on Slovakia, $(\Xi^{y_n}_{X_{nk}})_{SVK,AUT}$. In other words, we analyze if a shock on a regressor in Austria will affect outward FDI from Belgium to Slovakia. We observe a significant effect for the host GDP variable and bilateral distance but the effect is non-significant for the four other regressors. For instance, increasing the GDP of Austria by 1% will reduce outward FDI from Belgium to Slovakia by 0.147%. Finally, the last row of Table 9 studies significance of the difference between the indirect effect of Mexico on the United-States and the indirect effect of Canada on United-States ,$(\Xi^{y_n}_{X_{nk}})_{USA,MEX} - (\Xi^{y_n}_{X_{nk}})_{USA,CAN}$. We observe significant difference between those indirect effects for GDP and bilateral distance. In other words, the effect of a variation of Mexican GDP on outward FDI from Belgium to the United States will be statistically different from the effect of the same variation of Canadian GDP on outward FDI from Belgium to the United States.

Table 9: Inference on elements of impact matrices

	LGDP	*LPOP*	*OECD*	*LDIS*	*TARIFF*	*MP*
$\frac{1}{n}\operatorname{tr}(\Xi^{y_n}_{X_{nk}})$	1.113^{***}	-0.586^{**}	1.039^{*}	-1.202^{***}	0.106	1.213
	(0.242)	(0.255)	(0.613)	(0.200)	(0.083)	(1.190)
$(\Xi^{y_n}_{X_{nk}})_{SVK,AUT}$	-0.147^{*}	0.078	-0.137	0.160^{**}	-0.014	-0.157
	(0.084)	(0.058)	(0.120)	(0.075)	(0.013)	(0.226)
$(\Xi^{y_n}_{X_{nk}})_{USA,MEX} - (\Xi^{y_n}_{X_{nk}})_{USA,CAN}$	0.009^{*}	-0.005	0.008	-0.010^{**}	0.001	0.010
	(0.005)	(0.004)	(0.008)	(0.005)	(0.001)	(0.016)

Standard errors are between brackets; AUT stands for Austria, CAN for Canada, MEX for Mexico, SVK for Slovakia and USA for the United States; $\frac{1}{n}\operatorname{tr}(\Xi^{y_n}_{X_{nk}})$ is the average direct effect, $(\Xi^{y_n}_{X_{nk}})_{SVK,AUT}$ is the indirect effect between Austria and Slovakia; $(\Xi^{y_n}_{X_{nk}})_{USA,MEX} - (\Xi^{y_n}_{X_{nk}})_{USA,CAN}$ is the difference between the indirect effect of a change in x in Mexico on outward FDI in the United States and the indirect effect of a change in x in Canada on outward FDI in the United States; *, ** and *** correspond to significance at the 10%, 5% and 1% respectively.

6. Conclusions

This paper firstly develops the asymptotic theory of the matrix exponential spatial specification (MESS) in both the dependent variable and error terms. We show that the GMME is consistent and asymptotically normal even in the presence of unknown heteroskedasticity as long as the interaction matrix has zero diagonal elements. Besides we show that if the interaction matrices for the dependent variable and the error terms commute, the QMLE may also be consistent and asymptotically normal in the presence of unknown heteroskedasticity. In the homoskedastic case, we develop a best optimal GMME which is much simpler than the best optimal GMME for the SAR specification since moment conditions do not depend on estimated parameters. In case of non-normality, the homoskedastic best optimal GMME is shown to be more efficient than the QMLE. In the heteroskedastic case, a best optimal GMME cannot be derived except if we know the

structure of heteroskedasticity. We thus develop an optimal GMME which is shown to be more efficient than the QMLE. We also derive a lemma to perform inference on the elements, or functions of them, of the impact matrices implied by the reduced form of the MESS, which is very important for applied economists. Monte Carlo experiments are conducted and show the good small sample properties of the proposed estimators. Finally, we apply our estimators to show that outward FDI from Belgium are mainly characterized by the vertical type. We also compare SARAR and MESS(1,1) impacts and note that they are very similar, which pleads for the use of the latter. When the spatial process is stable, the MESS has many advantages over the SAR model.

Appendix A.

For the best GMME in the homoskedastic case, we show that adding any other moments to the selected ones cannot improve the asymptotic efficiency using the redundancy conditions in Breusch et al. (1999). Suppose that we have a set of moment conditions $\mathrm{E}[g_n^*(\gamma)] = 0$ with the corresponding optimal GMME being $\hat{\gamma}_n^*$. Adding some additional moment conditions $\mathrm{E}[g_n(\gamma)] = 0$ to $\mathrm{E}[g_n^*(\gamma)] = 0$, we have an optimal GMME $\hat{\gamma}_n$ using both sets of moment conditions. Then the moment conditions $\mathrm{E}[g_n(\gamma)] = 0$ are redundant given $\mathrm{E}[g_n^*(\gamma)] = 0$ if the asymptotic variances of $\hat{\gamma}_n^*$ and $\hat{\gamma}_n$ are the same. Let $V_n^* = n\,\mathrm{E}[g_n^*(\gamma_0)g_n^{*'}(\gamma_0)]$, $V_{n,21} = n\,\mathrm{E}[g_n(\gamma_0)g_n^{*'}(\gamma_0)]$, $G_n^* = \mathrm{E}\,\frac{\partial g_n^*(\gamma_0)}{\partial \gamma'}$ and $G_n = \mathrm{E}\,\frac{\partial g_n(\gamma_0)}{\partial \gamma'}$. The following two lemmas from Breusch et al. (1999) give conditions for moment redundancy.

Lemma 2. *The following statements are equivalent: (a)* $\mathrm{E}[g_n(\gamma)] = 0$ *is redundant given* $\mathrm{E}[g_n^*(\gamma)] = 0$*; (b)* $G_n = V_{n,21}V_n^{*-1}G_n^*$*; and (c) there exists a matrix* T *such that* $G_n^* = V_n^*T$ *and* $G_n = V_{n,21}T$*.*

Lemma 3. *Let the set of moment conditions to be considered be* $\mathrm{E}[g_n(\gamma)] = \mathrm{E}[g_{1n}'(\gamma), g_{2n}'(\gamma), g_{3n}'(\gamma)]' = 0$*, or simply* $g = (g_1', g_2', g_3')'$*. Then* $(g_2', g_3')'$ *is redundant given* g_1 *if and only if* g_2 *is redundant given* g_1 *and* g_3 *is redundant given* g_1*.*

Proof of Proposition 8. To show that $\hat{\gamma}_n^*$ is the best GMME within the class of GMMEs with linear and quadratic moments, we prove that the moment condition $\mathrm{E}[g_n(\gamma)] = 0$, where $g_n(\gamma)$ is a set of arbitrary linear and quadratic moments in (12), is redundant given the moment conditions $\mathrm{E}[g_n^*(\gamma)] = 0$. By Lemmas 2 and 3, it is sufficient to show that there exists a matrix T such that $G_n = \mathrm{E}\,\frac{\partial g_n(\gamma_0)}{\partial \gamma'} = V_{n,21}T$ and $G_n^* = V_n^*T$, where $V_{n,21} = n\,\mathrm{E}(g_n(\gamma_0)g_n^{*'}(\gamma_0)) = \frac{1}{n}\begin{pmatrix} \frac{1}{2}\sigma_0^4\omega_n'\omega_n^* & \mu_3\omega_{nd}^{'}F_n^* \\ \mu_3F_n^{'}\omega_{nd}^* & \sigma_0^2F_n'F_n^* \end{pmatrix} + \frac{(\mu_4-3\sigma_0^4)}{n}\begin{pmatrix} \omega_{nd}'\omega_{nd}^* & 0 \\ 0 & 0 \end{pmatrix}$, with $\omega_n = (\mathrm{vec}(P_{n1}^s), \ldots, \mathrm{vec}(P_{n,k_p}^s))$, $\omega_n^* = (\mathrm{vec}(P_{n1}^{*s}), \ldots, \mathrm{vec}(P_{n,k^*+4}^{*s}))$, $\omega_{nd} = (\mathrm{vec_D}(P_{n1}), \ldots, \mathrm{vec_D}(P_{n,k_p}))$ and $\omega_{nd}^* = (\mathrm{vec_D}(P_{n1}^*), \ldots, \mathrm{vec_D}(P_{n,k^*+4}^*))$, by Lemma A.2 in the supplementary file.

Let $P^*_{\alpha n} = P^*_{n1} - \frac{(\eta_4-3)-\eta_3^2}{(\eta_4-1)-\eta_3^2}P^*_{n2} - \frac{\sigma_0^{-1}\eta_3}{(\eta_4-1)-\eta_3^2}P^*_{n3}$, $P^*_{\tau n} = M_n$, $P^*_{\beta nl} = P^*_{n,l+4}$ for $l = 1,\ldots,k^*$, and $F^*_{\alpha n} = \frac{\eta_4-1}{(\eta_4-1)-\eta_3^2}F^*_{n2} - \frac{\eta_3^2}{(\eta_4-1)-\eta_3^2}F^*_{n3}(\frac{1}{n}l'_n e^{\tau_0 M_n}W_nX_n\beta_0) - \frac{2\sigma_0\eta_3}{(\eta_4-1)-\eta_3^2}F^*_{n4}$. If $e^{\tau_0 M_n}X_n$ does not contain an intercept term, let $F^*_{\beta n} = \frac{\eta_4-1}{(\eta_4-1)-\eta_3^2}F^*_{n1} - \frac{\eta_3^2}{(\eta_4-1)-\eta_3^2}F^*_{n3}(\frac{1}{n}l'_n e^{\tau_0 M_n}X^*_n)$; otherwise, let $F^*_{\beta n} = \frac{\eta_4-1}{(\eta_4-1)-\eta_3^2}F^*_{n1}(I_{k^*}, 0_{k^*\times 1}) + \frac{\eta_4-1}{(\eta_4-1)-\eta_3^2}c(\tau_0)F^*_{n3}e'_{kk} - \frac{\eta_3^2}{(\eta_4-1)-\eta_3^2}F^*_{n3}(\frac{1}{n}l'_n e^{\tau_0 M_n}X_n)$, where e_{kj} is the jth unit vector in R^k. Then

$$\epsilon'_n(\gamma)(P^*_{\alpha n}\epsilon_n(\gamma), P^*_{\tau n}\epsilon_n(\gamma), P^*_{\beta n1}\epsilon_n(\gamma), \ldots, P^*_{\beta nk^*}\epsilon_n(\gamma)) = \epsilon'_n(\gamma)(P^*_{n1}\epsilon_n(\gamma), \ldots, P^*_{n,k^*+4}\epsilon_n(\gamma))\Delta_P,$$

where

$$\Delta'_P = \begin{pmatrix} 1 & -\frac{(\eta_4-3)-\eta_3^2}{(\eta_4-1)-\eta_3^2} & -\frac{\sigma_0^{-1}\eta_3}{(\eta_4-1)-\eta_3^2} & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & I_{k^*} \end{pmatrix}.$$

If $e^{\eta_0 M_n}X_n$ does not contain an intercept term, $(F^*_{\alpha n}, F^*_{\beta n}) = (F^*_{n1}, F^*_{n2}, F^*_{n3}, F^*_{n4})\Delta_{F1}$, where

$$\Delta'_{F1} = \begin{pmatrix} 0 & \frac{\eta_4-1}{(\eta_4-1)-\eta_3^2} & -\frac{\eta_3^2}{(\eta_4-1)-\eta_3^2}(\frac{1}{n}l'_n e^{\tau_0 M_n}W_nX_n\beta_0) & -\frac{2\sigma_0\eta_3}{(\eta_4-1)-\eta_3^2} \\ \frac{\eta_4-1}{\eta_4-1-\eta_3^2}I_{k^*} & 0 & -\frac{\eta_3^2}{(\eta_4-1)-\eta_3^2}(\frac{1}{n}l'_n e^{\tau_0 M_n}X^*_n)' & 0 \end{pmatrix};$$

otherwise, $(F^*_{\alpha n}, F^*_{\beta n}) = (F^*_{n1}, F^*_{n2}, F^*_{n3}, F^*_{n4})\Delta_{F2}$, where

$$\Delta'_{F2} = \begin{pmatrix} 0 & \frac{\eta_4-1}{(\eta_4-1)-\eta_3^2} & -\frac{\eta_3^2}{(\eta_4-1)-\eta_3^2}(\frac{1}{n}l'_n e^{\tau_0 M_n}W_nX_n\beta_0) & -\frac{2\sigma_0\eta_3}{(\eta_4-1)-\eta_3^2} \\ \frac{\eta_4-1}{\eta_4-1-\eta_3^2}(I_{k^*}, 0_{k^*\times 1}) & 0 & \frac{\eta_4-1}{(\eta_4-1)-\eta_3^2}c(\tau_0)e_{kk} - \frac{\eta_3^2}{(\eta_4-1)-\eta_3^2}(\frac{1}{n}l'_n e^{\tau_0 M_n}X_n)' & 0 \end{pmatrix}.$$

Let $\Delta_{PF} = \begin{pmatrix} \Delta_P & 0 \\ 0 & \Delta_{F1} \end{pmatrix}$ if $e^{\eta_0 M_n}X_n$ does not contain an intercept term and $\Delta_{PF} = \begin{pmatrix} \Delta_P & 0 \\ 0 & \Delta_{F2} \end{pmatrix}$ otherwise. Then $g_n^{*'}(\gamma)\Delta_{PF} = \epsilon'_n(\gamma)(P^*_{\alpha n}\epsilon_n(\gamma), P^*_{\tau n}\epsilon_n(\gamma), P^*_{\beta n1}\epsilon_n(\gamma), \ldots, P^*_{\beta nk^*}\epsilon_n(\gamma), (F^*_{\alpha n}, F^*_{\beta n}))$. Let

$$\Delta'_T = \begin{pmatrix} \sigma_0^{-2} & 0 & 0 & (\sigma_0^{-2}, 0) \\ 0 & \sigma_0^{-2} & 0 & (0,0) \\ 0 & 0 & b' & (0, -\sigma_0^{-2}I_k) \end{pmatrix},$$

where $b' = (b'_1, \ldots, b'_{k^*})$ with $b_l = \frac{\sigma_0^{-3}\eta_3}{(\eta_4-1)-\eta_3^2}e'_{kl}$. Define $T = \Delta_{PF}\Delta_T$. We shall show that $G_n = V_{n,21}T$ and $G^*_n = V^*_n T$ for this T.

Let $J_n = I_n - \frac{1}{n}l_n l'_n$ and P_n be any $n\times n$ matrix with trace zero. The following identities are useful to show the desired results: (a) $\mathrm{vec_D}(P^*_{\alpha n}) = \frac{2}{(\eta_4-1)-\eta_3^2}\mathrm{vec_D}(\mathbb{W}_n) - \frac{\sigma_0^{-1}\eta_3}{(\eta_4-1)-\eta_3^2}J_n e^{\tau_0 M_n}W_nX_n\beta_0$; (b) $\mathrm{vec_D}(P^*_{\beta nl}) = J_n e^{\tau_0 M_n}X^*_{nl}$; (c) $\sum_{l=1}^{k^*}\mathrm{vec_D}(P^*_{\beta nl})e'_{kl} = J_n e^{\tau_0 M_n}X_n$; (d) $\sigma_0^2 F^*_{\alpha n} + \mu_3\,\mathrm{vec_D}(P^*_{\alpha n}) = \sigma_0^2 e^{\tau_0 M_n}W_nX_n\beta_0$; (e) $F^*_{\beta n} - \frac{\eta_3^2}{(\eta_4-1)-\eta_3^2}\sum_{l=1}^{k^*}\mathrm{vec_D}(P^*_{\beta nl})e'_{kl} = e^{\tau_0 M_n}X_n$; (f) $\mathrm{vec_D}'(P_n)F^*_{\beta n} = \frac{\eta_4-1}{(\eta_4-1)-\eta_3^2}\mathrm{vec_D}'(P_n)e^{\tau_0 M_n}X_n$; (g) $\mu_3\,\mathrm{vec_D}'(P_n)F^*_{\alpha n} + \sigma_0^4\,\mathrm{tr}(P^s_n P^*_{\alpha n}) + (\mu_4 - 3\sigma_0^4)\,\mathrm{vec_D}'(P_n)\,\mathrm{vec_D}(P^*_{\alpha n}) = \sigma_0^4\,\mathrm{tr}(P^s_n\mathbb{W}_n)$.

Since $g_n^{*'}(\gamma)\Delta_{PF} = \epsilon_n'(\gamma)(P_{\alpha n}^*\epsilon_n(\gamma), P_{\tau n}^*\epsilon_n(\gamma), P_{\beta n1}^*\epsilon_n(\gamma), \ldots, P_{\beta nk^*}^*\epsilon_n(\gamma), F_{\alpha n}^*, F_{\beta n}^*)$ as shown above and $P_{\beta nl}^*$'s are diagonal matrices, we have

$$V_{n,21}\Delta_{PF} = \mathrm{E}[g_n(\gamma_0)(g_n^{*'}(\gamma_0)\Delta_{PF})] = \frac{1}{n}\begin{pmatrix} \sigma_0^4\omega_n' \operatorname{vec}(P_{\alpha n}^*) & \sigma_0^4\omega_n' \operatorname{vec}(P_{\tau n}^*) & \sigma_0^4\omega_{nds}'\omega_{nd\beta}^* & \mu_3\omega_{nd}'(F_{\alpha n}^*, F_{\beta n}^*) \\ \mu_3 F_n' \operatorname{vec}_{\mathrm{D}}(P_{\alpha n}^*) & \mu_3 F_n' \operatorname{vec}_{\mathrm{D}}(P_{\tau n}^*) & \mu_3 F_n'\omega_{nd\beta}^* & \sigma_0^2 F_n'(F_{\alpha n}^*, F_{\beta n}^*) \end{pmatrix}$$
$$+ \frac{(\mu_4 - 3\sigma_0^4)}{n}\begin{pmatrix} \omega_{nd}' \operatorname{vec}_{\mathrm{D}}(P_{\alpha n}^*) & \omega_{nd}' \operatorname{vec}_{\mathrm{D}}(P_{\tau n}^*) & \omega_{nd}'\omega_{nd\beta}^* & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

where $\omega_{nd\beta}^* = (\operatorname{vec}_{\mathrm{D}}(P_{\beta n1}^*), \ldots, \operatorname{vec}_{\mathrm{D}}(P_{\beta nk^*}^*))$ and $\omega_{nds} = (\operatorname{vec}_{\mathrm{D}}(P_{n1}^s), \ldots, \operatorname{vec}_{\mathrm{D}}(P_{n,k_p}^s))$. The $V_{n,21}T = (V_{n,21}\Delta_{PF})\Delta_T$ is a 2×3 block matrix. By (g), the $(1,1)$th block of $V_{n,21}T$ is $\frac{1}{n}\sigma_0^2\omega_n' \operatorname{vec}(\mathbb{W}_n)$; the $(1,2)$th block is $\frac{1}{n}\sigma_0^2\omega_n' \operatorname{vec}(M_n)$; by (c) and (f), the $(1,3)$th block is 0; by (d), the $(2,1)$th block is $\frac{1}{n}F_n' e^{\tau_0 M_n} W_n X_n \beta_0$; the $(2,2)$th block is 0; by (e), the $(2,3)$th block is $-\frac{1}{n}F_n' e^{\tau_0 M_n} X_n$. Thus $V_{n,21}T = G_n$.

Furthermore, as $g_n^*(\gamma)$ is a special case of $g_n(\gamma)$, $G_n^* = V_n^* T$. Then $\Lambda_n^* = G_n^{*'} V_n^{*-1} G_n^* = G_n^{*'} T = (G_n^{*'}\Delta_{PF})\Delta_T = \mathrm{E}\,\frac{\partial(g_n^{*'}(\gamma_0)\Delta_{PF})}{\partial\gamma_0}\Delta_T$, which has the explicit expression in (15) by some computation. The asymptotic distribution of $\hat{\gamma}_n^*$ follows by Proposition 6. □

Acknowledgments

We would like to thank Yvan Stroppa and Matthias Pécot for providing valuable help with the Monte Carlo simulations, Peter Burridge and seminar participants of the economics department of the University of Murcia and University of Orléans.

References

Amemiya, T. (1985), *Advanced econometrics*, Harvard University Press, Cambridge, MA.

Baltagi, B. H., Egger, P. and Pfaffermayr, M. (2007), 'Estimating models of complex FDI: Are there third-country effects', *Journal of Econometrics* **140**, 260–281.

Baltagi, B. H., Egger, P. and Pfaffermayr, M. (2008), 'Estimating regional trade agreement effects on FDI in an interdependent world', *Journal of Econometrics* **145**, 194–208.

Blonigen, B. A., Davies, R. B., Waddell, G. R. and Naughton, H. T. (2007), 'FDI in space: Spatial autoregressive relationships in foreign direct investment', *European Economic review* **51**, 1303–1325.

Breusch, T., Qian, H., Schmidt, P. and Wyhowski, D. (1999), 'Redundancy of moment conditions', *Journal of Econometrics* **91**, 89–111.

Chiu, T. Y. M., Leonard, T. and Tsui, K.-W. (1996), 'The matrix-logarithmic covariance model', *Journal of the American Statistical Association* **91**(433), 198–210.

Coughlin, C. C. and Segev, E. (2000), 'Foreign direct investment in China: A spatial econometric study', *The World Economy* **23**, 1–23.

Davies, R. B. (2005), 'Fragmentation of headquarter services and FDI', *North American Journal of Economics and Finance* **16**, 61–79.

Debarsy, N. and Ertur, C. (2010), 'Testing for spatial autocorrelation in a fixed effects panel data model', *Regional Science and Urban Economics* **40**, 453–470.

Elhorst, J. P. (2010), 'Applied spatial econometrics: Raising the bar', *Spatial Economic Analysis* **5**(1), 9–28.

Elhorst, J. P., Lacombe, D. J. and Piras, G. (2012), 'On model specification and parameter space definitions in higher order spatial econometric models', *Regional Science and Urban Economics* **42**(1), 211–220.

Garretsen, H. and Peeters, J. (2009), 'FDI and the relevance of spatial linkages: do third country effects matter for Dutch FDI?', *Review of World Economics* **145**, 319–338.

Han, X. and Lee, L.-f. (2012), 'Model selection using j-test for the spatial autoregressive model vs. the matrix exponential spatial model', *Regional Science and Urban Economics* **Forthcoming**.

Kelejian, H. H. and Prucha, I. R. (1998), 'A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances', *Journal of Real Estate Finance and Economics* **17**, 99–121.

Kelejian, H. H. and Prucha, I. R. (2001), 'On the asymptotic distribution of the Moran I test statistic with applications', *Journal of Econometrics* **104**, 219–257.

Kelejian, H. H. and Prucha, I. R. (2010), 'Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances', *Journal of Econometrics* **157**, 53–67.

Lee, L.-F. (2003), 'Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances', *Econometric Reviews* **22**(4), 307–335.

Lee, L.-F. (2004), 'Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models', *Econometrica* **72**(6), 1899–1925.

Lee, L.-F. (2007), 'GMM and 2SLS estimation of mixed regressive, spatial autoregressive models', *Journal of Econometrics* **137**, 489–514.

Lee, L.-F. and Liu, X. (2010), 'Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances', *Econometric Theory* **26**, 187–230.

LeSage, J. P. and Pace, R. K. (2007), 'A matrix exponential spatial specification', *Journal of Econometrics* **140**, 190–214.

LeSage, J. and Pace, R. K. (2009), *Introduction to spatial econometrics*, Vol. 196, Chapman & Hall/CRC.

Lin, X. and Lee, L.-F. (2010), 'GMM estimation of spatial autoregressive models with unknown heteroskedasticity', *Journal of Econometrics* **157**, 34–52.

Liu, X., Lee, L.-F. and Bollinger, C. R. (2010), 'An efficient GMM estimator of spatial autoregressive models', *Journal of Econometrics* **159**, 303–319.

Smith, T. E. (2009), 'Estimation bias in spatial models with strongly connected weight matrices', *Geographical Analysis* **41**, 307–332.

White, H. (1980), 'A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity', *Econometrica* **48**, 817–838.