

Monte Carlo Simulation of The Multivariate Spatial Durbin Model for Complex Data Sets

Nur Atikah, Basuki Widodo, Mardlijah, Swasono Rahardjo,
Sri Harini, Riski Nur Istiqomah Dinnullah

Faculty of Science and Data Analytics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia

Faculty of Mathematics and Science
Universitas Negeri Malang
Malang, Indonesia

Faculty of Science and Technology
Universitas Islam Negeri Maulana Malik Ibrahim
Malang, Indonesia

Faculty of Science and Technology
Universitas PGRI Kanjuruhan Malang
Malang, Indonesia

email: nur.atikah.fmipa@um.ac.id,
b_widodo@matematika.its.ac.id, bwdantr@gmail.com

(Received July 29, 2024, Accepted August 24, 2024,
Published August 30, 2024)

Abstract

The Multivariate Spatial Durbin Model (MSDM) is a significant advance in spatial econometrics, very relevant in the context of research problems. This model extends spatial analysis by capturing the complexity and dynamism of interactions between variables in a spatial context that is often ignored by classical spatial models.

Key words and phrases: Multivariate Spatial Durbin Model, Monte Carlo Simulation, Maximum Likelihood Estimation, Maximum Likelihood Ratio Test, Complex Data Sets.

AMS (MOS) Subject Classifications: 62H12, 62H15.

The corresponding author is Basuki Widodo.

ISSN 1814-0432, 2025, <https://future-in-tech.net>

Furthermore, this article aims to estimate the parameters of MSDM model applied to large and complex data sets through Monte Carlo simulations. This model was then estimated using Maximum Likelihood Estimation (MLE), and to test the accuracy of the model using the Maximum Likelihood Ratio Test (MLRT) with a computational approach. The research results show that the MSDM model parameter estimates are accurate as indicated by an accuracy value that is smaller than the 5% significance level. The model becomes more efficient as the sample size increases.

1 Introduction

The Spatial Durbin Model (SDM) has gained significant popularity among researchers, particularly in univariate response studies. Numerous empirical investigations utilizing SDM with univariate responses have been conducted by ([1],[2], [3], [4], [5]). However, the increasing complexity of real-world problems and the need for more comprehensive modeling approaches have driven the development of spatial regression models with multivariate responses [6].

This article meticulously develops the Multivariate Spatial Durbin Model (MSDM) as an extension of the MSAR by incorporating endogenous and exogenous spatial interactions between cross-sectional units. The model extends spatial analysis by capturing the complexity and dynamism of interactions between variables in a spatial context that classical spatial models often ignore. The MSDM model is estimated using Maximum Likelihood Estimation (MLE), and the accuracy of the model is rigorously tested using the Maximum Likelihood Ratio Test (MLRT) with a computational approach. This paper's theoretical and computational findings are substantiated by a comprehensive simulation study using Monte Carlo simulation. Simulation results show high agreement with theoretical expectations, thus confirming the validity and reliability of the research findings [7].

This MSDM model estimation is developed from the highly reliable MLE algorithm. This MLE can be used as an alternative in spatial data analysis, where MLE is effective in solving large data cases [8] and is reliable in improving the estimation results of geostatistical models [9].

2 Multivariate Spatial Durbin Model

Incorporating the methodological framework of the Multivariate Spatial Autoregressive (MSAR) model as delineated in [10] and integrating the spatial dependencies among exogenous variables, we extend this foundation to formulate the Multivariate Spatial Durbin Model (MSDM). This advanced model is presented as follows:

$$Y_h = \rho_{hh} WY_h + \sum_{h' \neq h}^p \rho_{h'h} WY_{h'} + \sum_{k=1}^q b_{kh} X_k + \sum_{k=1}^q W X_k \theta_{kj} + \varepsilon_h \quad (2.1)$$

Y_h represents the h -th column vector of the dependent variable matrix Y formed from N locations, with $1 \leq h \leq p$, X_k denotes the matrix of independent variables. The term WY_h signifies the influence of other dependent variables, with the corresponding parameter $\rho_{h'h}$ referred to as the extra activity effect, also known as the endogenous effect.

2.1 Parameter Estimation of MSDM

Equation (2.1) can be reformulated in matrix notation as follows:

$$Y = WYP + XB + WX\Theta + E \quad (2.2)$$

where

$$Y = (Y_1, Y_2, \dots, Y_p)$$

$$X = (X_1, X_2, \dots, X_q)$$

$$P = (\rho_{h'h}) \in R^{p \times p}$$

$$B = (b_{kh}) \in R^{q \times p}$$

$$\Theta = (\theta_{kj}) \in R^{q \times p}$$

$$E = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p) \in R^{N \times p}$$

The MSDM model (2.2) can be reformulated as an MSAR model by defining the matrix $Z = [I \ X \ WX]^T$ and the vector $\tilde{B} = [B \ \Theta]$, thereby simplifying equation (2.2) to: $Y = (I \otimes PW)Y + \tilde{Z}\beta + \varepsilon$, where

$$\varepsilon = (I - I \otimes PW)Y - \tilde{Z}\beta, \quad (2.3)$$

with $Y = \text{vec}(Y) = (Y_1^T, Y_2^T, \dots, Y_p^T)^T \in R^{N \times p}$

$$\tilde{Z} = I_p \otimes Z$$

$$\varepsilon \sim N(0, \sigma^2 I)$$

Using the derivative of equation (2.3), the log-likelihood function is obtained as follows.

$$\begin{aligned} \ln(L(P, \beta, \sigma^2|Y)) &= -\frac{p}{2} \ln(2\pi) - \frac{p}{2} \ln(\sigma^2) + \ln|I - I \otimes PW| \\ &\quad - \frac{1}{2\sigma^2} \left((I - I \otimes PW)Y - \tilde{Z}\beta \right)^T \\ &\quad - \frac{1}{2\sigma^2} \left((I - I \otimes PW)Y - \tilde{Z}\beta \right) \end{aligned} \quad (2.4)$$

2.1.1 Estimation of Parameter β on MSDM Model

The parameter β estimation can be achieved by maximizing the log-likelihood function in equation (2.4). This involves differentiating equation (2.4) with respect to β ($\frac{\partial \ln(L)}{\partial \beta} = 0$). Consequently, the estimator for the parameter β is given by:

$$\hat{\beta} = \left(\tilde{Z}^T \tilde{Z} \right)^{-1} \tilde{Z}^T (I - I \otimes PW) Y \quad (2.5)$$

Proposition 2.1. *If β represents the parameter matrix of the MSDM model, then $\hat{\beta}$ serves as the estimator parameter at the i -th location, which is defined as follows:*

$$\hat{\beta} = \left(\tilde{Z}^T \tilde{Z} \right)^{-1} \tilde{Z}^T (I - I \otimes PW) Y$$

Proof.

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_q \end{bmatrix} = \begin{bmatrix} \left(\tilde{Z}^T \tilde{Z} \right)^{-1} \tilde{Z}^T (I - I \otimes PW) Y_1 \\ \left(\tilde{Z}^T \tilde{Z} \right)^{-1} \tilde{Z}^T (I - I \otimes PW) Y_2 \\ \dots \\ \left(\tilde{Z}^T \tilde{Z} \right)^{-1} \tilde{Z}^T (I - I \otimes PW) Y_q \end{bmatrix} = \left(\tilde{Z}^T \tilde{Z} \right)^{-1} \tilde{Z}^T (I - I \otimes PW) Y_{(n \times q)}$$

Proposition 2.2. *If $\hat{\beta}$ represents the estimator parameter of the MSDM model, then $\hat{\beta}$ is an unbiased estimator for β .*

Proof.

The unbiased nature of the $\hat{\beta}$ estimator will be demonstrated by showing that $E(\hat{\beta}) = \beta$

$$\begin{aligned} E(\hat{\beta}) &= E \left(\left(\tilde{Z}^T \tilde{Z} \right)^{-1} \tilde{Z}^T (I - I \otimes PW) Y \right) \\ E(\hat{\beta}) &= E \left(\left(\tilde{Z}^T \tilde{Z} \right)^{-1} \tilde{Z}^T (I - I \otimes PW) (I - I \otimes PW)^{-1} (\tilde{Z}\beta + \epsilon) \right) \end{aligned}$$

$$E(\hat{\beta}) = E\left(\left(\tilde{Z}^T \tilde{Z}\right)^{-1} \tilde{Z}^T \tilde{Z} \beta\right)$$

$$E(\hat{\beta}) = \beta$$

Proposition 2.3. *If $\hat{\beta}$ is the estimator parameter of the MSDM model, then $\hat{\beta}$ constitutes an efficient estimator of β .*

Proof.

The efficient property of the $\hat{\beta}$ estimator can be demonstrated by showing that the variance of the minimum $\hat{\beta}$ is minimized.

$$Var(\hat{\beta}) = E\left(\left(\hat{\beta} - E(\hat{\beta})\right)\left(\hat{\beta} - E(\hat{\beta})\right)^T\right) = \left(Z^T \tilde{Z}\right)^{-1} \sigma^2$$

Let $\hat{\beta}^*$ be another linear estimator of $\hat{\beta}$, then

$$\hat{\beta}^* = \left(\left(\tilde{Z}^T \tilde{Z}\right)^{-1} \tilde{Z}^T + c\right) (I - I \otimes PW) Y$$

$$\hat{\beta}^* = \left(\left(\tilde{Z}^T \tilde{Z}\right)^{-1} \tilde{Z}^T + c\right) (\tilde{Z} \beta + \epsilon)$$

$$\hat{\beta}^* = \beta + \left(\tilde{Z}^T \tilde{Z}\right)^{-1} \epsilon + c \tilde{Z} \beta + c \epsilon$$

where c is a constant matrix.

So $E(\hat{\beta}^*) = \beta + c \tilde{Z} \beta$.

The estimator $\hat{\beta}^*$ is assumed to be an unbiased estimator of $\hat{\beta}$, so that $E(\hat{\beta}^*) = \beta$. Then we get that $c \tilde{Z} \beta$ is a 0 matrix. So that,

$$\hat{\beta}^* - \beta = \left(\left(\tilde{Z}^T \tilde{Z}\right)^{-1} \tilde{Z}^T + c\right) \epsilon$$

$$Var(\hat{\beta}^*) = E\left(\left(\hat{\beta}^* - E(\hat{\beta}^*)\right)\left(\hat{\beta}^* - E(\hat{\beta}^*)\right)^T\right)$$

$$Var(\hat{\beta}^*) = \left(\left(\tilde{Z}^T \tilde{Z}\right)^{-1} \tilde{Z}^T + c\right) E(\epsilon \epsilon^T) \left(\tilde{Z} \left(\tilde{Z}^T \tilde{Z}\right)^{-1} + c^T\right)$$

$$Var(\hat{\beta}^*) = \sigma^2 \left(\left(\tilde{Z}^T \tilde{Z}\right)^{-1} \tilde{Z}^T + c\right) \left(\tilde{Z} \left(\tilde{Z}^T \tilde{Z}\right)^{-1} + c^T\right)$$

$$Var(\hat{\beta}^*) = \sigma^2 \left(\left(\tilde{Z}^T \tilde{Z}\right)^{-1} + c c^T\right)$$

$$Var(\hat{\beta}^*) = Var(\hat{\beta}) + \sigma^2 c c^T$$

So we get $Var(\hat{\beta}^*) - Var(\hat{\beta}) = H a \sigma^2 c c^T$.

Because $c c^T$ is a matrix whose elements are all non-negative and $\hat{\sigma}^2 \geq 0$,

then $\hat{\sigma}^2 cc^T \geq 0$. So that $Var(\hat{\beta}) \leq Var(\hat{\beta}^*)$. So, it can be concluded that $\hat{\beta}$ is efficient or has minimum variance.

2.1.2 Estimation of Parameter σ^2 in MSDM Model

The estimation of σ^2 is performed by maximizing the log-likelihood function presented in Equation (2.4), which involves differentiating Equation (2.4) with respect to σ^2 ($\frac{\partial \ln(L)}{\partial \sigma^2} = 0$). Hence, the estimator parameter for σ^2 is derived as follows:

$$\sigma^2 = \frac{1}{p}((I - I \otimes PW)Y - \tilde{Z}\beta)^T((I - I \otimes PW)Y - \tilde{Z}\beta) \quad (2.6)$$

Proposition 2.4. *If $\hat{\sigma}^2$ is the estimator parameter from the MSDM model, then $\sigma^2 = \frac{SSE}{n-2 \text{tr}(S)+\text{tr}(S^T S)}$ is an unbiased estimator for σ^2 .*

Proof.

The unbiased property of the estimator $\hat{\sigma}^2$ will be demonstrated by showing that $E(\hat{\sigma}^2) = \sigma^2$ where

$$\begin{aligned} S &= \left(\tilde{Z}^T Z\right)^{-1} \tilde{Z}^T (I - I \otimes PW) \\ E(\hat{\sigma}^2) &= E\left(\frac{1}{p}\left((I - I \otimes PW)Y - \tilde{Z}\beta\right)^T\left((I - I \otimes PW)Y - \tilde{Z}\beta\right)\right) \\ E(\hat{\sigma}^2) &= \frac{1}{p}E(\epsilon^T \epsilon) \\ E(\hat{\sigma}^2) &= \frac{1}{p}E(SSE) \\ E(SSE) &= E(\epsilon^T \epsilon) \\ E(SSE) &= E\left(\epsilon^T (I - S)^T (I - S) \epsilon\right) \\ E(SSE) &= E\left(\text{tr}\left(\epsilon^T (I - S)^T (I - S) \epsilon\right)\right) \\ E(SSE) &= (n - 2\text{tr}(S) + \text{tr}(S^T S)) \sigma^2 \\ SSE &= (n - 2\text{tr}(S) + \text{tr}(S^T S)) \sigma^2 \\ \sigma^2 &= \frac{SSE}{n-2 \text{tr}(S)+\text{tr}(S^T S)} \end{aligned}$$

So, it is proven that if $\hat{\sigma}^2$ is the estimator parameter of the MSDM model, then $\sigma^2 = \frac{SSE}{n-2 \text{tr}(S)+\text{tr}(S^T S)}$ is an unbiased estimator for σ^2 .

Corollary 2.5. *If $\hat{\Sigma}$ is an unbiased estimator for the variance-covariance matrix Σ , then $\hat{\Sigma}$ is also an unbiased estimator of the variance-covariance matrix Σ .*

Proof.

It will be demonstrated that $\hat{\Sigma} = \frac{Y^T(I-S)^T(I-S)Y}{\left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2}\right)}$

$$\begin{aligned} E(SSE_h) &= (n - 2tr(S) + tr(S^T S)) \sigma_h^2 \\ E(SE_h E_h) &= (n - 2tr(S) + tr(S^T S)) \sigma_{hh}^2 \\ E(\hat{\sigma}_h^2) &= \sigma_h^2 \\ E(\hat{\sigma}_{hh}^2) &= \sigma_{hh}^2 \\ \hat{\Sigma} &= \frac{Y^T(I-S)^T(I-S)Y}{\left(\frac{\hat{\sigma}_1^2}{\hat{\sigma}_2}\right)} \end{aligned}$$

So, it is proven that If $\hat{\Sigma}$ is an unbiased estimator for the variance-covariance matrix Σ , then $\hat{\Sigma}$ is also an unbiased estimator of the variance-covariance matrix Σ .

2.1.3 Estimation of Parameter P in MSDM Model

Based on equation (2.5), the estimation of P is determined by maximizing σ^2 in equation (2.6) by differentiating equation (2.5) with respect to ρ_0 and ρ_1 ($\frac{\partial(\sigma^2)}{\partial\rho_0} = 0$ and $\frac{\partial(\sigma^2)}{\partial\rho_1} = 0$). Thus, we obtain

$$\hat{\rho} = \rho_0 = \rho_1 = (Z^T Z)^{-1} Z^T Y \tag{2.7}$$

Proposition 2.6. *If \hat{P} is an estimator parameter from the MSDM model, then $\hat{\rho} = (Z^T Z)^{-1} Z^T Y$ is an unbiased estimator for P.*

Proof.

The unbiased nature of the estimator \hat{P} will be demonstrated by showing that $E(\hat{P}) = P$.

$$\begin{aligned} E(\hat{P}) &= E\left((Z^T Z)^{-1} Z^T Y\right) \\ E(\hat{P}) &= E\left((Z^T Z)^{-1} Z^T (I - (I \otimes PW))^{-1} (Z\beta + \epsilon)\right) \\ E(\hat{P}) &= E(P) \\ E(\hat{P}) &= P \end{aligned}$$

It is proven that If \hat{P} is an estimator parameter from the MSDM model, then $\hat{\rho} = (Z^T Z)^{-1} Z^T Y$ is an unbiased estimator for P.

Table 1: Estimation of the slope coefficients.

Slope	8 cn							
	50		100		300		500	
	Mean	St.Dev	Mean	St.Dev	Mean	St.Dev	Mean	St.Dev
$\rho_1 : 0.825$	0.813	0.074	0.821	0.065	0.151	0.066	0.821	0.068
$\rho_2 : 0.769$	0.750	0.078	0.753	0.077	0.759	0.084	0.764	0.085
$\rho_3 : 0.444$	0.432	0.132	0.443	0.158	0.434	0.155	0.440	0.157
$\beta_{11} : -0.017$	-0.022	0.016	-0.017	0.018	-0.016	0.018	-0.016	0.018
$\beta_{12} : -0.065$	-0.071	0.116	-0.056	0.110	-0.070	0.108	-0.064	0.113
$\beta_{13} : -0.047$	-0.047	0.097	-0.049	0.103	-0.049	0.113	-0.042	0.108
$\beta_{21} : -0.150$	-0.155	0.064	-0.154	0.070	-0.154	0.079	-0.151	0.073
$\beta_{22} : -0.090$	0.001	0.496	-0.084	0.449	-0.107	0.462	-0.070	0.485
$\beta_{23} : 1.318$	1.330	0.428	1.304	0.487	1.308	0.432	1.305	0.435
$\beta_{31} : 0.039$	0.058	0.081	0.043	0.084	0.039	0.089	0.037	0.090
$\beta_{32} : -0.646$	-0.517	0.542	-0.569	0.633	-0.684	0.553	-0.668	0.621
$\beta_{33} : -0.953$	-0.921	0.586	-0.950	0.575	-0.981	0.577	-0.959	0.529
$\theta_{11} : 0.0289$	0.0360	0.033	0.032	0.036	0.031	0.035	0.028	0.039
$\theta_{12} : 0.060$	0.059	0.225	0.062	0.195	0.074	0.199	0.050	0.195
$\theta_{13} : -0.519$	-0.522	0.197	-0.546	0.250	-0.517	0.229	-0.511	0.234
$\theta_{21} : -0.082$	-0.097	0.135	-0.092	0.155	-0.087	0.160	-0.084	0.154
$\theta_{22} : 0.034$	0.096	1.064	0.006	0.864	0.091	0.831	0.011	0.821
$\theta_{23} : -0.397$	-0.405	0.906	-0.538	1.018	-0.419	0.992	-0.315	0.975
$\theta_{31} : 0.068$	0.092	0.176	0.083	0.210	0.077	0.210	0.067	0.192
$\theta_{32} : -1.226$	-1.199	1.017	-1.246	1.102	-1.282	1.012	-1.201	1.122
$\theta_{33} : 0.039$	0.249	1.218	0.165	1.338	0.034	1.294	0.038	1.280

3 Results and Discussion

This section simulates the Multivariate Spatial Durbin Model (MSDM) estimations using limited samples with sizes $N = 50, 100, 300, 500$. The parameters ρ, β , and θ are set according to the values in Table 1. The simulation results indicate that the average parameter estimates increasingly converge toward their true values as the sample size increases and the standard deviation decreases. This signifies that the proposed estimation procedure performs effectively.

Table 2: MSDM Model Simulation Test Statistics.

Parameter	4cn			
	50	100	300	500
β_{11}	0.04113	0.04079	0.04702	0.04003
β_{12}	0.05474	0.05087	0.04543	0.04181
β_{13}	0.05387	0.04963	0.04317	0.04126
β_{21}	0.05002	0.04771	0.04395	0.04147
β_{22}	0.05007	0.05399	0.04982	0.04781
β_{23}	0.05704	0.05200	0.05017	0.04766
β_{31}	0.05810	0.05014	0.04870	0.04238
β_{32}	0.04932	0.04537	0.04948	0.04201
β_{33}	0.04859	0.04698	0.04465	0.04192
θ_{11}	0.05681	0.04981	0.05457	0.04287
θ_{12}	0.05041	0.04896	0.04553	0.04328
θ_{13}	0.05339	0.04981	0.04375	0.04576
θ_{21}	0.04834	0.04795	0.04536	0.04015
θ_{22}	0.04984	0.04745	0.04360	0.04442
θ_{23}	0.04734	0.04854	0.04456	0.04692
θ_{31}	0.05238	0.05008	0.04863	0.04684
θ_{32}	0.05137	0.05351	0.04955	0.04515
θ_{33}	0.04915	0.04774	0.05040	0.04477

Furthermore, statistical tests will be carried out for the MSDM model simulation, with the results presented in Table 2. The tests show that the values of each parameter are consistently at the 5% significance level, it is indicating that the MSDM model estimation procedure is statistically significant.

4 Conclusion

This article introduces a novel model, the Multivariate Spatial Durbin Model (MSDM), which extends the Multivariate Spatial Auto Regressive (MSAR) model by incorporating spatial effects on exogenous variables. MSDM model parameters are estimated using the Maximum Likelihood Estimator (MLE) method, while hypothesis testing is conducted using the Maximum Likelihood Ratio Test (MLRT) test statistic. Monte Carlo simulations demonstrate that the MSDM model parameter estimation utilizing the MLE method yields precise results. This finding is substantiated by the outcomes of the MLRT statistical test, which effectively maintains a 5% significance level, it is indicating the robustness of the model. Notably, the efficiency of the MSDM model exhibits a positive correlation with increasing sample size, suggesting its suitability for large datasets.

References

- [1] Y. Liu, H. Geng, Regional Competition in China under the Price Distortion of Construction Land: A Study Based on a Two-regime Spatial Durbin Model, *China and World Economy*, **27**, no. 4, (2019), 104–126.
- [2] J. Li, S. Li, Energy investment, economic growth and carbon emissions in China—Empirical analysis based on spatial Durbin model, *Energy Policy*, **140**, (2020), 2020111425.
- [3] H. Xie, Z. Ouyang, Y. Choi, Characteristics and influencing factors of green finance development in the Yangtze River delta of China: Analysis based on the spatial Durbin model, *Sustainability (Switzerland)*, **12**, no. 22, (2020), 1–15.

- [4] N. Atikah, S. Rahardjo, D. L. Afifah, N. Kholifia, Modeling Spatial Spillovers of regional economic growth in East Java: An empirical analysis based on Spatial Durbin Model, *Journal of Physics: Conference Series*, 1872, no. 1, (2021).
- [5] H. Xiao, J. Mao, Effects of postgraduate education on technological innovation: a study based on the spatial Durbin model, *Asia Pacific Education Review*, **22**, no. 1, (2021), 89–99.
- [6] A. M. Astuti, Setiawan, I. Zain, J. D. T. Purnomo, A modified generalized estimating equation approach for simultaneous spatial Durbin panel model: A case study of economic growth in ASEAN countries, *Decision Science Letters*, **12**, no. 2, (2023), 369–388.
- [7] E. Zio, *The Monte Carlo Simulation Method for System Reliability and Risk Analysis*, London: Springer, 2012.
- [8] S. Abdulah, H. Ltaief, Y. Sun, M. Genton, D. Keyes, Parallel Approximation of the Maximum Likelihood Estimation for the Prediction of Large-Scale Geostatistics Simulations, *IEEE International Conference on Cluster Computing (CLUSTER)*, (2018), 98–108.
- [9] E. Pardo-Igúzquiza, Maximum Likelihood Estimation of Spatial Covariance Parameters, *Mathematical Geology*, **30**, (1998), 95–108.
- [10] X. Zhu, D. Huang, R. Pan, H. Wang, Multivariate spatial autoregressive model for large-scale social networks, *Journal of Econometrics*, **215**, no. 2, (2019), 591–606.