THE IMPACT OF FOREIGN DIRECT INVESTMENT ON INCOME INEQUALITY IN VIETNAM: A SPATIAL REGRESSION APPROACH

Ho Thuy Tien[1], Trinh Thi Lien[2*]

[1]University of Finance - Marketing, Vietnam
[2]Ho Chi Minh City Industry and Trade College, Vietnam
[*]Corresponding author: Email: ttlien@hitu.edu.vn

Received: August 04, 2025 *Accepted:* October 22, 2025 *Published:* November 25, 2025
DOI: 10.52932/jfmr.v3i5ene.1057

Appendix 1. Practical realities and the existing gaps

From a theoretical perspective, FDI is regarded as a driver for modernizing the economies of developing countries, typically through creating more jobs, integrating advanced technologies into operations, developing new products, and supplementing domestic financial resources (Lall & Narula, 2004; Balasubramanyam, 2002). However, an unbalanced allocation of FDI can lead to economic disparities, exacerbating income gaps between different social groups or regions. The concentration of investment capital in only certain industries or areas may leave other regions lagging behind, thereby deepening income inequality (Neagu et al., 2016; Lin et al., 2013; Wu & Hsu, 2012).

According to the Dependence Theory, FDI is a tool that developed countries use to create growth momentum for developing countries through the transfer of technological advances and capital (Cardoso, 1982; Frank, 1967). However, this process is often only concerned with industries where the intensity of capital use is higher than that of labor, leading to a reduction in the demand for untrained human resources and thereby aggravating the income inequality in the society of the receiving country.

In contrast to Dependence Theory, Modernization Theory argues that the technological and skill spillovers generated by FDI increase the demand for highly skilled labor (Mihalache-O'Keef & Li, 2011). When this process of knowledge and technical transfer occurs evenly throughout the domestic market, it helps adjust the mismatch between labor supply and demand, thereby reducing income inequality compared to the initial stage. At this point, the FDI-driven economic growth model is often likened to an inverted-U Kuznets curve, which reflects the non-linear relationship between FDI and income inequality.

In practice, Vietnam has recently achieved significant results in terms of GDP growth. However, alongside this development is a rise in income inequality among different social groups (Nguyen Thi Thai Hung & Nguyen Quynh Tho, 2019; Duong et al., 2017). Uneven growth poses the risk of trapping the country in the middle-income trap, which is a major obstacle to the transformation and development of emerging economies. If income inequality continues to widen, it could trigger profound negative social impacts, necessitating costly policy interventions. For example, the 2011 uprising in Egypt is a vivid case of the severe consequences that can arise when income inequality exceeds public tolerance. In this context, Vietnam needs to promptly recognize this as an urgent issue that deserves proper attention. The World Bank's 2018 report revealed that 76% of urban residents and 53% of rural residents expressed concern about income inequality, and this concern is even more pronounced among the youth— a group expected to play a proactive role in society in the near future.

From an empirical perspective, analyses of the correlation between FDI and income inequality in Vietnam remain scattered and show no clear consensus. Some studies suggest that FDI contributes to narrowing income inequality, whereas others reveal the opposite trend (Do et al., 2024; Phan, 2022; Ho Dinh Bao et al., 2020; Duong et al., 2017). In addition, a non-linear relationship between the two variables has also been documented (Nguyen Thi Thai Hung & Nguyen Quynh Tho, 2019), indicating that the impact of FDI is not fixed but can vary depending on the stage of development and the unique socio-economic characteristics of each region.

Furthermore, according to Wei et al. (2009), sectors that attract large amounts of FDI tend to boost production and business activities, thereby drawing labor from less developed areas. This labor migration may widen income inequality between regions with and without FDI, demonstrating that the spatial impact of FDI on income inequality is not confined to the host locality but can also extend to neighboring areas.

Appendix 2. Spatial weight matrix

Symbol:

$$W=\begin{bmatrix} W_{11} & \cdots & W_{1n} \\ \vdots & \ddots & \vdots \\ W_{n1} & \cdots & W_{nm} \end{bmatrix}$$

The spatial weight matrix (W) is a square matrix of size (n × n), where n is the number of spatial units in the model. An important characteristic of this matrix is symmetry. If unit i has a relationship with unit j, the influence from i to j is equal to the influence from j to i ($W_{ij} = W_{ji}$).

The values on the main diagonal are zero: Since a locality cannot have a spatial relationship with itself, all elements on the main diagonal are set to zero ($W_{ii} = 0$ for all i).

Based on the method of defining spatial linkages between localities, the W matrix can be categorized as follows:

- First-order contiguity spatial Weight Matrix:

According to Coughlin & Segev (2000), two geographic units are considered spatially connected if they share a common border. In this case, if locality i shares a boundary with locality j (i ≠ j), then a direct spatial relationship exists between them. Consequently, the elements in the spatial weight matrix (W) are determined according to the following principle:

$$W_{ij} = \begin{cases} 1 & i,j \text{ share a common border} \\ 0 & i,j \text{ do not share a common border} \end{cases}$$

- Distance-based Spatial Weight Matrix:

According to Anselin (2013), a spatial weight matrix can be constructed based on the geographic distance between observation units. Let d be the distance threshold and d_{ij} be the actual distance between localities i and j. In this case, the elements of the spatial weight matrix W are determined according to the following rule:

$$W_{ij} = \begin{cases} 1, 0 \leq d_{ij} \leq d \\ 0, d > d_{ij} \end{cases}$$

- Inverse Distance-Based Matrix:

The spatial weight matrix can also be constructed based on the inverse distance between localities. Accordingly, the weight between two localities i and j is determined using the following formula:

$$W_{ij} = \begin{cases} \dfrac{1}{d_{ij}^{\alpha}}, & i \neq j \\ 0, & i = j \end{cases}$$

where: d_{ij} is the actual distance between localities i and j.

α is the exponent that determines the rate at which influence decays with distance, typically taking the values α=1 or α=2.

In practice, Anselin (2013) proposes a row-standardization of the spatial weight matrix to ensure that the sum of the weights in each row equals 1. The standardized weight matrix is computed as:

$$W_{ij}^{s} = \frac{W_{ij}}{\sum_j W_{ij}}$$

With this standardization, the sum of the elements in each row of the spatial weight matrix equals 1, thereby ensuring that the aggregate influence from neighboring localities on a given locality is always constrained within the range of 0 to 1.

3.2.3. Spatial Regression Model for Panel Data

The spatial regression model applied to panel data can be expressed in the following general form:

$$y_{i,t} = \alpha + \rho \sum_{j=1}^{n} W_{ij} y_{ij} + \sum_{j=1}^{n} x_{ij}\beta_k + \sum_{k=1}^{k} \sum_{j=1}^{n} W_{ij} x_{ijt}\theta_k + \tau_i + \xi_t + \vartheta_{it} \quad (2.1)$$

Với $\vartheta_{it} = \lambda \sum_{j=1}^{n} m_{ij}\vartheta_k + \varepsilon_{t,t}$ (2.2)

$\varepsilon \sim (\sigma^2 l)$ (2.3)

i = 1, 2…n; t= 1,2…t

y: dependent variable

x: independent variable

W: spatial weight matrix

If $\theta = 0$ the model (2.1) becomes the Spatial Autocorrelation Model (SAC)

If $\lambda = 0$ the model (2.1) becomes the Spatial Durbin Model (SDM).

If $\theta = 0$ và $\lambda = 0$ the model (2.1) becomes the Spatial Autoregressive Model (SAR)

If $\theta = 0$ và $\rho = 0$ the model (2.1) becomes the Spatial Error Model (SEM)

If $\theta = 0$, $\rho = 0$ và $\tau_i = \Psi \sum_{j=1}^{n} W_{ij}\tau_i + \varsigma_i$ the model (2.1) becomes the Generalized Spatial Panel Random Effects Model (GSPRE)

Appendix 3. Spatial model selection tests

Testing the selection between SAR and SDM models (Likelihood Ratio – LR test)

The Likelihood Ratio (LR) test is used to compare the SDM with the SAR to determine whether the SDM can be reduced to the SAR model. Specifically, the test is constructed based on the following hypotheses:

- Null Hypothesis (H_0): θ = 0 (The SAR model is more appropriate).
- Alternative Hypothesis (H_1): θ ≠ 0 (the SDM model is more appropriate).

When the p-value is less than 5%, we reject the null hypothesis (H_0). This indicates that the SDM model is more suitable, as it cannot be simplified to the SAR model.

Testing the selection between SEM and SDM Models (Likelihood Ratio – LR Test)

The Likelihood Ratio (LR) test can also be applied to compare the SDM with the SEM to determine whether the SDM can be simplified to the SEM model. The hypotheses for this test are as follows:

- Null Hypothesis (H_0): $\theta + \rho\beta = 0$ (The SEM model is more appropriate).
- Alternative Hypothesis (H_1): $\theta + \rho\beta \neq 0$ (the SDM model is more appropriate).

If the p-value is less than 5%, we reject the null hypothesis (H_0), which means that the SDM model is a more appropriate choice than the SEM model.

Model selection test among SEM, SAC, and GSPRE

To determine the best regression model among the three models SEM, SAC and GSPRE, the author uses information criteria such as AIC (Akaike Information Criterion) developed by Hirotsugu Akaike in 1973 (Akaike,

1973) and BIC (Bayesian Information Criterion) developed by Gideon E. Schwarz in 1978 (Schwarz, 1978). Both of these criteria are used to evaluate the fit of the regression model to the data. The smaller the value of these indices, the better the model fits the data.

The formulas for AIC and BIC are as follows:

AIC: $AIC=2k-2\ln(L)$

BIC: $BIC=\ln(n)k-2\ln(L)$

Where: k is the number of parameters in the regression model, L is the likelihood function, n is the number of observations in the dataset.

Appendix 4. VIF Values of Variables in the Model

Variable	VIF	1/VIF
PE	2.46	0.4511
PI	1.93	0.5523
TC	1.56	0.5914
IQ	1.65	0.7711
URB	1.18	0.7653
PINV	1.65	0.8421
POP	1.22	0.8782
FDI	1.18	0.8614
UNE	1.11	0.9739
Mean VIF	1.52	

Appendix 5. Regression Results

Main	SEM_RE	SEM_FE	SAR_RE	SAR_FE	SDM_RE	SDM_FE	SAC_FE	GSPRE
FDI	-0.00086*** [-0.41]	-0.000264*** [0.10]	-0.000674*** [-0.44]	-0.000407*** [-0.14]	0.000224 [0.08]	0.000216 [0.07]	-0.000207** [0.07]	-0.00064*** [-0.40]
URB	2.42E-06 [0.67]	1.78E-06 [0.04]	0.00000247 [0.74]	0.0000174 [0.26]	0.00000426 [0.67]	-0.00000842 [-0.14]	6.46E-06 [0.26]	2.28E-06 [1.48]
POP	-0.00044 [-1.46]	-0.00184*** [-4.22]	-0.000440* [-1.67]	-0.00164*** [-4.24]	-0.000406** [-1.64]	-0.00186*** [-4.21]	-0.00181*** [-11.60]	-0.00048*** [-1.48]
UNE	0.00044** [-0.21]	0.00062** [-0.41]	-0.000477* [-0.18]	0.000201 [0.10]	-0.00044 [-0.26]	-0.000447 [-0.24]	-0.000666* [-1.84]	-0.00048 [-0.44]
IQ	0.0126** [1.40]	0.00672** [-0.40]	0.0114*** [1.47]	0.00648*** [1.04]	0.00162*** [0.14]	-0.0107 [-0.76]	-0.00672 [-0.40]	0.0126** [1.44]
PI	-0.00885** [-2.51]	-0.0147*** [-5.17]	-0.00777*** [-2.77]	-0.0172*** [-4.24]	-0.00758** [-2.01]	-0.0152*** [-2.84]	-0.0147*** [-5.78]	-0.00702** [-2.47]
PE	0.00105*** [0.47]	0.00107*** [0.57]	0.00154 [0.75]	-0.000281 [-0.07]	-0.000827 [-0.55]	0.00188 [0.72]	0.00104 [0.58]	0.00104** [0.41]
TC	-0.00012 [-0.75]	-0.00072 [-1.05]	-0.000157 [-0.84]	-0.0000125 [-0.02]	-0.00011 [-0.77]	-0.000725 [-0.87]	-0.00072 [-1.07]	-0.00015* [-0.81]
PINV	-0.0174 [-1.12]	-0.0705*** [-2.87]	-0.0171 [-1.24]	-0.0875*** [-5.57]	-0.0255 [-1.45]	-0.0787*** [-2.77]	-0.0722*** [-4.18]	-0.0175 [-1.51]
_cons	0.417*** [71.85]		0.507*** [15.70]		0.542*** [14.88]			0.417*** [57.77]
Spatial								
lambda	0.258*** [5.45]	-0.0522*** [-0.58]					-0.0777 [-1.04]	0.257*** [7.74]
rho			0.272*** [5.77]	0.242*** [5.50]	0.252*** [4.78]	-0.0224 [-0.41]	0.0781 [0.72]	
Variance								
sigma2_e	0.0000700***	0.0000548***	0.0000777***	0.0000718***	0.0000787***	0.0000557***	0.0000701***	

Main	**SEM_RE**	**SEM_FE**	**SAR_RE**	**SAR_FE**	**SDM_RE**	**SDM_FE**	**SAC_FE**	**GSPRE**
	[17.50]	[18.71]	[17.47]	[18.50]	[17.47]	[18.71]	[24.87]	
ln_phi	-4.550477 [-2.75]							
sigma_mu								0.000712 [1.54]
sigma_e								0.00857*** [41.51]
Hausman Test	47.79***		77.25***		132.72***			
LR Test				111.52***				
		12.23						
AIC		-4897.14					-4610.73	-4724.73
BIC		-4834.58					-4460.81	-4591.39
	SEM_RE	**SEM_FE**	**SAR_RE**	**SAR_FE**	**SDM_RE**	**SDM_FE**	**SAC_FE**	**GSPRE**
		Main		**LR_Direct**		**LR_Direct**	**LR_Direct**	**Main**
FDI		-0.000276*** [0.10]		-0.000502** [-0.10]		0.000545 [0.11]	0.00051 [0.11]	-0.00065*** [-0.50]
URB		0.00000197 [0.05]		0.0000155 [0.26]		-0.0000118 [-0.21]	0.00000551 [0.26]	2.28E-06 [1.48]
POP		-0.00191*** [-5.22]		-0.00161*** [-5.28]		-0.00185*** [-5.25]	-0.00180*** [-12.55]	-0.00048*** [-1.58]
UNE		0.000776** [-0.51]		0.000185 [0.06]		-0.000445 [-0.25]	-0.000804* [-1.64]	-0.00048 [-0.55]
IQ		0.00684** [-0.50]		0.00685 [1.08]		-0.0108 [-0.81]	-0.00564 [-0.56]	0.0126** [1.55]
PI		-0.01966*** [-5.18]		-0.0165*** [-4.26]		-0.0128*** [-2.88]	-0.0145*** [-5.84]	-0.00602** [-2.46]
PE		0.00115*** [0.56]		-0.000521 [-0.10]		0.00185 [0.58]	0.00108 [0.56]	0.00104** [0.41]

Main	SEM_RE	SEM_FE	SAR_RE	SAR_FE	SDM_RE	SDM_FE	SAC_FE	GSPRE
TC		-0.00091 [-1.05]		-0.0000264 [-0.04]		-0.000646 [-0.68]	-0.000855 [-1.12]	-0.00015* [-0.81]
PINV		-0.0798*** [-2.86]		-0.0864*** [-5.41]		-0.0662*** [-2.85]	-0.0808*** [-4.16]	-0.0185 [-1.51]
LR_Indirect								
FDI				-0.0000684** [-0.10]		-0.00565** [-0.68]	-0.0000445 [-0.11]	
URB				0.00000454 [0.25]		0.000166** [2.14]	0.00000151 [0.42]	
POP				-0.000561** [-2.46]		0.000558 [0.40]	-0.000146 [-0.85]	
UNE				0.0000465 [0.08]		-0.0026 [-0.81]	-0.000088 [-0.68]	
IQ				0.00261 [1.02]		0.00685 [0.56]	0.0000655 [0.05]	
PI				-0.00586*** [-5.18]		-0.0166* [-1.62]	-0.00126 [-0.84]	
PE				-0.000111 [-0.11]		0.00241 [0.56]	0.000118 [0.28]	
TC				-0.00000608 [-0.04]		0.000552 [0.42]	-0.000044 [-0.58]	
PINV				-0.0264*** [-2.66]		-0.0625* [-1.85]	-0.00642 [-0.84]	
LR_Total								
FDI				-0.0004** [-0.10]		-0.00526** [-0.58]	-0.000265** [0.08]	
URB				0.0000166 [0.26]		0.000188* [1.84]	0.00000682 [0.26]	

Main	**SEM_RE**	**SEM_FE**	**SAR_RE**	**SAR_FE**	**SDM_RE**	**SDM_FE**	**SAC_FE**	**GSPRE**
POP				-0.00250*** [-5.18]		-0.00148 [-1.55]	-0.00165*** [-8.15]	
UNE				0.000254 [0.06]		-0.00554 [-0.85]	-0.000882* [-1.86]	
IQ				0.0128 [1.08]		-0.000615 [-0.05]	-0.00588 [-0.52]	
PI				-0.0251*** [-4.26]		-0.0268*** [-5.15]	-0.0158*** [-5.26]	
PE				-0.000452 [-0.10]		0.00424 [0.61]	0.00118 [0.56]	
TC				-0.0000585 [-0.04]		-0.000064 [-0.06]	-0.000888 [-1.10]	
PINV				-0.115*** [-5.58]		-0.156*** [-2.88]	-0.0882*** [-5.51]	
N	882	882	882	882	882	882	882	882