Urbanization and Income Inequality in Thailand

Acumsiri Tantiakrnpanit

***Abstract*—**This paper aims to examine the relationship between urbanization and income inequality in Thailand during the period 2002–2020, using a panel of data for 76 provinces collected from Thailand's National Statistical Office (Labor Force Survey: LFS), as well as geospatial data from the U.S. Air Force Defense Meteorological Satellite Program (DMSP) and the Visible Infrared Imaging Radiometer Suite Day/Night band (VIIRS-DNB) satellite for 19 selected years. This paper employs two different definitions to identify urban areas: 1) Urban areas defined by Thailand's National Statistical Office (LFS), and 2) Urban areas estimated using nighttime light data from the DMSP and VIIRS-DNB satellite. The second method includes two sub-categories: 2.1) Determining urban areas by calculating nighttime light density with a population density of 300 people per square kilometer, and 2.2) Calculating urban areas based on nighttime light density corresponding to a population density of 1,500 people per square kilometer. The empirical analysis based on Ordinary Least Squares (OLS), fixed effects, and random effects models reveals a consistent U-shaped relationship between income inequality and urbanization. The findings from the econometric analysis demonstrate that urbanization or population density has a significant and negative impact on income inequality. Moreover, the square of urbanization shows a statistically significant positive impact on income inequality. Additionally, there is a negative association between logarithmically transformed income and income inequality. This paper also proposes the inclusion of satellite imagery, geospatial data, and spatial econometric techniques in future studies to conduct quantitative analysis of spatial relationships.

***Keywords*—**Income inequality, nighttime light, population density, Thailand, urbanization.

I. Introduction

In the past, most people worldwide resided in small communities. However, over the past few centuries, particularly in recent decades, there has been a significant shift in living patterns. This transformation occurred due to a mass migration of populations from rural areas to urban areas starting in 2007. As a result, the urban population eventually surpassed the rural population, with 3.35 million people residing in urban areas compared to 3.33 million in rural areas. In the case of Thailand, this transition took place in 2018, when the number of people living in urban areas (34.68 million) became comparable to the rural population (34.75 million). Subsequently, Thailand experienced a continuous increase in urban populations, surpassing the rural population [1]. The influx of immigrants from rural to urban areas is driven by the combination of agglomeration economies and the presence of urban mechanisms that support development. This combination allows producers to exploit labor efficiently rather than relying solely on economies of scale. This notion aligns with the findings of [2], which suggests that urbanization can contribute to reducing national inequality by bridging the gap between urban and rural areas through various mechanisms.

The concept of urbanization has experienced a progressive transformation over the years, with a particular focus on creating a balanced city system. Additionally, different approaches to urban development have diverse effects on economic growth and equality across societies. Moreover, according to [3], a sound urban structure can potentially influence the distribution of income.

Factors related to economic development and specific areas of interest contribute to the complexity of understanding such relationships, making it challenging to draw definitive conclusions. Consequently, we conducted an empirical analysis in Thailand, utilizing provincial data from the LFS published by the National Statistical Office (NSO). Additionally, we incorporated alternative geospatial data from the U.S. Air Force DMSP and the VIIRS-DNB satellite. The objective of this paper is twofold. Firstly, we aimed to aggregate province-level data using conventional statistics from Thailand's NSO as the primary data sources, covering the period from 2002 to 2020. Additionally, alternative geospatial data were integrated and employed in this study. Secondly, study aims to quantitatively investigate the relationship between urbanization and income inequality in Thailand.

II. Review of Literature

A. Studies Urbanization Effects

Urbanization influences the growth process, the efficiency of growth, the extent of income inequality, external scale economies, knowledge spillovers, driving the spatial evolution of production and population agglomeration.

B. Urbanization and Increased Income Inequality

Reference [10] studied relationship between urbanization and income inequality in China. It was found that the threshold rate of urbanization in this result was 0.53, meaning implicitly stated that the province had an inverted-U relationship. The threshold of urbanization of more than 0.53 will experience reduced income inequality. And the study found that rich provinces tend to have lower rural-urban inequality and a higher inflow of immigrants, and the rural-urban wage gap makes significant contributions to income inequality. This is consistent with the findings of [4], who found that the income inequality increased in post-reform China can be causally associated with urbanization. Examination of time series data demonstrate that urbanization exerts a nonlinear impact on income inequality, with a lagged effect. This influence leads to an initial decrease in the Gini index, followed by a subsequent

Acumsiri Tantikarnpanit is with Thammasat University, Thailand (e-mail: acumsiri.tan@st.econ.tu.ac.th).

increase. Furthermore, outcomes derived from diverse models provided empirical evidence of a noteworthy and affirmative correlation between urbanization and income inequality in sub-Saharan Africa [5]. In addition, [6] found that occupations in large cities also offered higher wages compared to occupations that did not require larger cities. It also found the fact that the wages of some occupation groups increased with the size of the cities.

Meanwhile, [7] studied the impact of the changing dual economic structure in Asia specifically in four Asian countries including India, the Peoples' Republic of China, Indonesia, and the Philippines. According to the Kuznets model, which posits a bell-shaped relationship between income inequality and income levels, utilizing Theil's method to calculate income inequality would be a suitable metric in this study. Their results found that the observed increase in inequality is due to changes in the dual economic structure and how urbanization changes depending on the country. In addition, [8] found that urbanization positively affected on the urban-rural income gap (URIG).

C. Urbanization and Reduction in Income Inequality

Reference [9] explored the impact of urbanization on income inequality in Vietnam. From the model, it was found that the impact of urbanization has a negative regression with a statistically significant 1% in the model implying the negative impact of urbanization on income inequality in Vietnamese provinces. This means that higher urbanization contributes to reducing income inequality. And they found the urbanization affects income inequality in the long term but not in the short term. They used the square of urbanization in the model to test the nonlinear relationship between urbanization and income inequality. It was found that urbanization increased income inequality in the early stages until reaching a certain threshold. After that, income inequality will be reduced. This is consistent with the finding of [7] who studied the impact of the changing dual economic structure in Asia specifically in four Asian countries. Based on the Kuznets model and inequality measurement using the inequality index (Theil's Method), their results found that in the People's Republic of China, a decrease in urbanization contributed to a 2.3% reduction in inequality, making it the only country with such a decrease. Moreover, they found the income gap between urban and rural areas has narrowed slightly in Indonesia and the Philippines. Similarly, [10] is interested in studying income inequality in developing countries, particularly least developed countries (LDCs). They found the mechanisms behind urbanization can help reduce national inequality by narrowing the gap between urban and rural areas as: 1. improving per capita income of rural areas 2. improving rural income and 3. increasing rural income.

III. Data

A. LFS

An official LFS is regularly conducted by the Thailand's NSO. Microeconomic data and macroeconomic data are in quarterly data segments from the period of 2002 to 2020. We collected income distribution data for 76 provinces since 2002 to 2011 and collected income distribution data for 77 provinces since 2012, representatives of provinces in each region.

B. DMSP NTL Satellite Imageries

In 1972, the US Air Force's DMSP initiated the use of remote sensing technology to capture luminance in the visible and near infrared light spectrum emitted by urban areas. Initially, this program was primarily focused on weather forecasting. Publicly available annual data for 1992-2013 are accessible, with a spatial resolution of 30 × 30 arc seconds, equivalent to approximately 1 × 1 kilometer. Each pixel within the data represents the intensity of nighttime lights (NTL), ranging from 0 to 63, where 0 represents complete darkness, while 63 represents maximum brightness. In this study, the annual DMSP NTL satellite imageries in 2002 to 2012 were converted into a provincial NTL index that represents average NTL density derived from a population density of 300 people and 1,500 people per square kilometer to define urban areas.

C. VIIRS-DNB NTL Satellite Imageries

The Visible Infrared Imaging Radiometer Suite (VIIRS) Day/Night Band (DNB) initiated the use of remote sensing technology to capture luminance in the visible and near infrared light spectrum emitted by urban areas. Initially, the program primarily concentrated on monitoring and examining changes and characteristics on a global and regional level. Annual data for 2011 are accessible to the public, collected over a swath width of 3,060 kilometers. The data collection involves 22 distinct spectral bands within the electromagnetic spectrum, covering wavelengths ranging from 0.412 μm to 12.01 μm. Each pixel within the data reflects the intensity of NTL. In this study, the annual VIIRS-DNB NTL satellite imageries in 2013 to 2020 were converted into a provincial NTL index that represents average NTL density derived from a population density of 300 people and 1,500 people per square kilometer to define urban areas.

IV. Methodology

A. Construction of the Econometrically Test

To econometrically test the relationship between urbanization and income inequality in municipalities in each province measured by Gini coefficient, we run some reduced-form regression, like those in (1)-(5):

$$Gini_{it} = \beta_1 U_{it-1} + \beta_2 U_{it-1}^2 + \mu_{it} \tag{1}$$

$$Gini_{it} = \beta_1 U_{it-1} + \beta_2 U_{it-1}^2 + \beta_3 lnGPP_{it-1} + \alpha_t + \mu_{it} \tag{2}$$

$$Gini_{it} = \beta_1 U_{it-1} + \beta_2 U_{it-1}^2 + \beta_3 lnGPP_{it-1} + \alpha_t + \gamma_r + \mu_{it} \tag{3}$$

$$Gini_{it} = \beta_1 U_{it-1} + \beta_2 U_{it-1}^2 + \beta_3 lnGPP_{it-1} + \alpha_t + \gamma_r + \varphi_i + \mu_{it} \tag{4}$$

$$Gini_{it} = \beta_1 U_{it-1} + \beta_2 U_{it-1}^2 + \beta_3 lnGPP_{it-1+\beta_4 UlnGPP_{it-1}+\alpha_t+\gamma_r+\varphi_i+\mu_{it}} \tag{5}$$

where $Gini_{it}$ is income inequality in municipalities in province i in time t measured by the Gini coefficient, U_{it-1} is the urbanization rate which is the proportion of population in municipalities in province i versus the population in province i in time $t-1$, U^2_{it-1} is square of U_{it-1}. $lnGPP_{it-1}$ is log of GPP per capita in municipalities in province i in time $t-1$, $UlnGPP_{it-1}$ is the product term of urbanization rate and log of GPP per capita in municipalities in province i in time $t-1$, α_t is year fixed effect in time t, γ_r is region fixed effect in time r and φ_i is province random effect in province i.

To econometrically test the relationship between urbanization and income inequality in municipalities in each province measured by the Atkinson index, we run some reduced-form regression, like those in (6)-(10):

$$Atkinson_{it} = \beta_1 U_{it-1} + \beta_2 U^2_{it-1} + \mu_{it} \quad (6)$$

$$Atkinson_{it} = \beta_1 U_{it-1} + \beta_2 U^2_{it-1} + \beta_3 lnGPP_{it-1} + \alpha_{t+\mu_{it}} \quad (7)$$

$$Atkinson_{it} = \beta_1 U_{it-1} + \beta_2 U^2_{it-1} + \beta_3 lnGPP_{it-1} + \alpha_{t+\gamma_r+\mu_{it}} \quad (8)$$

$$Atkinson_{it} = \beta_1 U_{it-1} + \beta_2 U^2_{it-1} + \beta_3 lnGPP_{it-1} + \alpha_t + \gamma_r + \varphi_i + \mu_{it} \quad (9)$$

$$Atkinson_{it} = \beta_1 U_{it-1} + \beta_2 U^2_{it-1} + \beta_3 lnGPP_{it-1} + \beta_4 UlnGPP_{it-1} + \alpha_t + \gamma_r + \varphi_i + \mu_{it} \quad (10)$$

where $Atkinson_{it}$ is income inequality in municipalities in province i in time t measured by the Atkinson index, U_{it-1} is the urbanization rate which is the proportion of population in municipalities in province i versus the population in province i in time $t-1$, U^2_{it-1} is square of U_{it-1}. $lnGPP_{it-1}$ is log of GPP per capita in municipalities in province i in time $t-1$, $UlnGPP_{it-1}$ is product term of urbanization rate and log of GPP per capita in municipalities in province i in time $t-1$, α_t is year fixed effect in time t, γ_r is region fixed effect in time r and φ_i is province random effect in province i.

B. Construction of the Urban Areas Based on Nighttime Light Density

The average of nighttime light is computed as:

$$AN = \frac{\sum_{t=1}^{19} N_t * t}{number\ of\ the\ year} \quad (11)$$

where AN is the average of nighttime light, N_t is the nighttime light of each province in time t, t is time from 1 to 19 years and represents 2002 to 2020 and the $number\ of\ the\ year$ is equal to 19.

The nighttime light density is computed as:

$$\mathrm{ND} = \frac{AN}{PA^2} \quad (12)$$

where ND is the nighttime light density, AN is the average of nighttime light, and PA^2 is the pixel of area each province.

The average of population is computed as:

$$AP = \frac{\sum_{t=1}^{19} P_t * t}{number\ of\ the\ year} \quad (13)$$

where AP is the average of population, P_t is the population of each province in time t, t is time from 1 to 19 years and represents 2002 to 2020, and $number\ of\ the\ year$ is equal to 19.

The population density is computed as:

$$PD = \frac{AP}{A^2} \quad (14)$$

where PD is the population density, AP is the average of population, and A^2 is area of province (square kilometer).

The equation to find the value of nighttime light density to find the urban area is computed as:

$$ND_{it} = \beta_1 + \beta_2 PD_{it} + \mu_{it} \quad (15)$$

where PD is the population density equal 300 people per square kilometer and 1,500 people per square kilometer. We will get (16) and (17) for urban areas based on nighttime light density derived from DMSP NTL satellite imageries and VIIRS-DNB NTL satellite imageries, respectively.

C. Fixed Effect Models and Random Effect Models

The basic model of panel data consists of fixed effect model and random effect model. A fixed effect examines if the intercept varies across group or time, while a random effects model explores difference in the error variance component across individuals or time. The core difference between fixed and random is the role of dummy variables. For fixed effect, a parameter estimated of dummy variables is a part of the intercept whereas random effect is error components. The functional forms of one-way fixed effect models and random effect models are shown below.

Fixed Effect Model:

$$y_{it} = \alpha_i + \beta_1 x_{1it} + .. + \beta_k x_{kit} + u_{it} \quad (16)$$

where α_i is cross sectional units and constant value, indicating the intercept for each unit.

Random Effect Model:

$$y_{it} = \alpha_i + \beta_1 x_{1it} + .. + \beta_k x_{kit} + u_{it} \quad (17)$$

where α_i is cross sectional units but not constant value, indicating the intercept was causally state. Therefore, the intercept α_i is rewritten as follows:

$$\alpha_i = \alpha_i + \varepsilon_i \quad (18)$$

Therefore, random effect models can be rewritten as follows:
Random Effect Model:

$$y_{it} = \alpha_i + \beta_1 x_{1it} + .. + \beta_k x_{kit} + \varepsilon_i + u_{it} \quad (19)$$

and error term (u_{it}) is independent identically distributed (i.i.d.) ($u_{it} \sim iid(0, \sigma^2_u)$). Fixed effect model examines individual differences in intercepts, assuming constant variance across the sample and same slope. The random effect model estimates error variance specific to individuals or time, assuming that the individual-specific effect is not correlated with the independent variables. Individual differences show specific errors of individuals, not in intercepts [11].

The choice between a fixed effect model and a random effect model in panel data analysis depends on which model is more appropriate for the data. The fixed effect model can be tested using an F-test, while the random effect model can be tested using the Lagrange multiplier (LM) test, developed by Breusch & Pagan (1979) [12]. Then, we compare OLS and fixed effect model that shows how much the fixed effect model can improve the goodness-of-fit, whereas the latter contrasts OLS with random effect model. Hausman test examines the similarity between fixed effect and random effect estimators.

F-test for Fixed Effect Models: F-test will compare between unrestricted models and restricted models.

$$F - test = \frac{(SSR_R - SSR_{UR})}{SSR_{UR}/(N-k-1)} \tag{20}$$

If the null hypothesis is rejected, it indicates a significant fixed effect or a significant increase in goodness-of-fit in the fixed effect model. Therefore, the fixed effect model is more appropriate than the pooled OLS.

Breusch-Pagan Lagrange Multiplier for Random Effect Models: If we suspect that heteroskedasticity depends on only independent variables, we can use the Breusch-Pagan test. We simply regress the square of error (u^2) to whatever independent variable we choose and run the appropriate F or LM test.

D. Income Inequality Indices

The Gini coefficient, developed by the Italian statistician Corrado Gini in 1912, measures income distribution across a population [13]. It often is a gauge of economic inequality, measuring income distribution, or less commonly, wealth distribution among a population. The Gini coefficient ranges from 0 to 1, where 0 indicates complete equality and 1 indicates complete inequality. But, the Gini coefficient could theoretically be greater than 1 because of the income or wealth of a population is negative. The Gini index is usually expressed through the Lorenz curve, which represents the distribution of income, through the relationship between the X-axis, representing the percentage of the cumulative population shares, and the Y-axis, representing the percentage of the cumulative income shares. The Gini coefficient is defined as:

$$Gini = 1 - \sum_{i=1}^{N}(x_i - x_{i-1})(y_i + y_{i-1}) \tag{21}$$

where i is individuals, N is the total number of intervals, x_i is point on the X-axis, and y_i is a point on the Y-axis. The Gini coefficient has a disadvantage that it is not additive across groups.

The Atkinson index, developed by the British economist Anthony Barnes Atkinson, measures income distribution [14]. The Atkinson index helps to determine which end of the distribution contributed most to the observed inequality. It can be turned into a normative measure by imposing a coefficient to weight incomes. This measurement is more sensitive to changes at the lower end of the income distribution if ε increases. That means it is less sensitive to changes at the lower end of the income distribution if ε approaches 0. The Atkinson index is defined as:

$$A_\varepsilon = 1 - \frac{x_e}{\bar{x}} = 1 - [\frac{1}{n}\sum_{i=1}^{n}[\frac{y_i}{\bar{x}}]^{1-\varepsilon}]^{\frac{1}{1-\varepsilon}} \tag{22}$$

where $\boldsymbol{\varepsilon}$ parameter is often called the "inequality aversion parameter".

V. Results

Tables I-III provide the relationship between urbanization and income inequality in municipalities in each province measured by the Gini coefficient calculated as:

$$Gini_{it} = \beta_1 U_{it-1} + \beta_2 U_{it-1}^2 + \mu_{it} \tag{23}$$

$$Gini_{it} = \beta_1 U_{it-1} + \beta_2 U_{it-1}^2 + \beta_3 lnGPP_{it-1} + \alpha_t + \mu_{it} \tag{24}$$

$$Gini_{it} = \beta_1 U_{it-1} + \beta_2 U_{it-1}^2 + \beta_3 lnGPP_{it-1} + \alpha_t + \gamma_r + \mu_{it} \tag{25}$$

$$U_{it-1} + \beta_2 U_{it-1}^2 + \beta_3 lnGPP_{it-1} + \alpha_t + \gamma + \varphi_i + \mu_{it_r} \tag{26}$$

$$Gini_{it} = \beta_1 U_{it-1} + \beta_2 U_{it-1}^2 + \beta_3 lnGPP_{it-1+\beta_4 UlnGPP_{it-1}+\alpha_t+\gamma_r+\varphi_i+\mu_{it}} \tag{27}$$

The results presented in Tables I-III show a negative and highly significant association between income (log of GPP per capita in municipalities in each province) and income inequality (Gini Coefficient in municipalities in each province). The significance of this association holds, excluding the urbanization rate when controlling for income at the province level and year fixed effects (24). Notably, the significance of the coefficient for the urbanization rate and the square of the urbanization rate also holds when controlling for region fixed effects (25). The coefficient for the urbanization rate remains significant even when introducing province-specific random effects (26). To capture the role of income levels, in (27) of Tables I-III, the interaction between the urbanization rate and province income is considered (interaction between the urbanization rate in municipalities in each province and the logarithm of GPP per capita in municipalities in each province). While the coefficient for the urbanization rate and income levels is both negative and significant, the coefficient for the square of the urbanization rate and the interaction between the two is positive, except for Table I.

Tables IV-VI provide the relationship between urbanization and income inequality in municipalities in each province measured by the Atkinson index calculated as:

$$Atkinson_{it} = \beta_1 U_{it-1} + \beta_2 U_{it-1}^2 + \mu_{it} \tag{28}$$

$$Atkinson_{it} = \beta_1 U_{it-1} + \beta_2 U_{it-1}^2 + \beta_3 lnGPP_{it-1} + \alpha_t + \mu_{it} \tag{29}$$

$$Atkinson_{it} = \beta_1 U_{it-1} + \beta_2 U_{it-1}^2 + \beta_3 lnGPP_{it-1} + \alpha_t + \gamma_r + \mu_{it} \tag{30}$$

$$Atkinson_{it} = \beta_1 U_{it-1} + \beta_2 U_{it-1}^2 + \beta_3 lnGPP_{it-1} + \alpha_t + \gamma_r + \varphi_i + \mu_{it} \quad (31)$$

$$Atkinson_{it} = \beta_1 U_{it-1} + \beta_2 U_{it-1}^2 + \beta_3 lnGPP_{it-1} + \beta_4 UlnGPP_{it-1} + \alpha_t + \gamma_r + \varphi_i + \mu_{it} \quad (32)$$

TABLE I
GINI COEFFICIENT IN MUNICIPALITIES IN PROVINCE (URBAN AREAS AS DEFINED BY THAILAND'S NSO (LFS))

Result of Equation	(23)	(24)	(25)	(26)	(27)
Dependent variable: Income Inequality (Gini Coefficient in municipalities in each province					
Urbanization rate	0.0028 ** (0.0014)	-0.0037 *** (0.0007)	-0.0031 *** (0.0008)	-0.0028 *** (0.0012)	-0.0030 *** (0.0013)
Square of Urbanization rate	-0.00002 ** (0.0000)	0.00003 *** (0.0000)	0.00003 *** (0.0000)	0.00003 *** (0.0000)	0.00003 *** (0.0000)
Log (GPP per capita)		-0.0405 *** (0.0051)	-0.0348 *** (0.0067)	-0.0738 *** (0.0092)	-0.0698 *** (0.0110)
*Urbanization rate * Log (GPP per capita)*					-0.0002 *** (0.0003)
Year FE	No	Yes	Yes	Yes	Yes
Region FE	No	No	Yes	Yes	Yes
Province (Random) Effects	No	No	No	Yes	Yes
Observations	1,386	1,377	1, 377	1, 377	1, 377
No. of provinces	77	77	77	77	77
R-Square	0.0636	0.6503	0.6780	0.4030	0.4049

Note: The time span is from 2002 to 2020; ***, **, and * indicate significance at 1%, 5%, and 10%, respectively.

TABLE II
GINI COEFFICIENT IN MUNICIPALITIES IN PROVINCE (CALCULATING URBAN AREAS BASED ON NIGHTTIME LIGHT DENSITY DERIVED FROM A POPULATION DENSITY OF 300 PEOPLE PER SQUARE KILOMETER)

Result of Equation	(23)	(24)	(25)	(26)	(27)
Dependent variable: Income Inequality (Gini Coefficient in municipalities in each province					
Urbanization rate	-0.0024 *** (0.0045)	-0.0018 *** (0.0007)	-0.0008 (0.0006)	-0.0007 ** (0.0003)	-0.0008 *** (0.0003)
Square of Urbanization rate	0.00002 *** (0.0000)	0.00001 *** (0.0000)	0.00001 (0.0000)	0.00001 (0.0000)	0.00000 (0.0000)
Log (GPP per capita)		-0.0327 *** (0.0083)	-0.0312 *** (0.0085)	-0.0502 *** (0.0092)	-0.0655 *** (0.0107)
*Urbanization rate * Log (GPP per capita)*					0.0005 ** (0.0002)
Year FE	No	Yes	Yes	Yes	Yes
Region FE	No	No	Yes	Yes	Yes
Province (Random) Effects	No	No	No	Yes	Yes
Observations	1,368	1,368	1,368	1,368	1,368
No. of provinces	76	76	76	76	76
R-Square	0.1113	0.7369	0.7659	0.1645	0.1656

Note: The time span is from 2002 to 2020; ***, **, and * indicate significance at 1%, 5%, and 10%, respectively.

TABLE III
GINI COEFFICIENT IN MUNICIPALITIES IN PROVINCE (CALCULATING URBAN AREAS BASED ON NIGHTTIME LIGHT DENSITY DERIVED FROM A POPULATION DENSITY OF 1,500 PEOPLE PER SQUARE KILOMETER)

Result of Equation	(23)	(24)	(25)	(26)	(27)
Dependent variable: Income Inequality (Gini Coefficient in municipalities in each province					
Urbanization rate	-0.0052 *** (0.0008)	-0.0037 *** (0.0009)	-0.0028 *** (0.0010)	-0.0018 * (0.0010)	-0.0025 ** (0.0011)
Square of Urbanization rate	0.00006 *** (0.0000)	0.00004 *** (0.0000)	0.00003 *** (0.0000)	0.00003 * (0.0000)	0.00002 ** (0.0000)
Log (GPP per capita)		-0.0328 *** (0.0087)	-0.0277 *** (0.0084)	-0.0450 *** (0.0093)	-0.0507 *** (0.0093)
*Urbanization rate * Log (GPP per capita)*					0.0007 *** (0.0003)
Year FE	No	Yes	Yes	Yes	Yes
Region FE	No	No	Yes	Yes	Yes
Province (Random) Effects	No	No	No	Yes	Yes
Observations	1,368	1,368	1,368	1,368	1,368
No. of provinces	76	76	76	76	76
R-Square	0.1266	0.7509	0.7855	0.1815	0.1838

Note: The time span is from 2002 to 2020; ***, **, and * indicate significance at 1%, 5%, and 10%, respectively.

Tables IV-VI present the results which show a negative and highly significant association between income (log of GPP per capita in municipalities in each province) and income inequality (Atkinson index in municipalities in each province). The significance of this association holds even when excluding the urbanization rate, when controlling for income at the province level and year fixed effects (29). Furthermore, the significance of the coefficient for the urbanization rate and the square of the urbanization rate also holds when controlling for region fixed effects (30). The coefficient for the urbanization rate remains significant even when introducing province-specific random effects (31). To capture the role of income levels, in (32) of Tables IV-VI, the interaction between the urbanization rate and provincial income is considered (interaction between the urbanization rate in municipalities in each province and the logarithm of GPP per capita in municipalities in each province). While the coefficients for the urbanization rate and income levels are both negative and significant, the coefficient for the square of the urbanization rate and the interaction between the two are both positive, except Table IV.

TABLE IV
ATKINSON INDEX IN MUNICIPALITIES IN PROVINCE (URBAN AREAS AS DEFINED BY THAILAND'S NSO (LFS))

Result of Equation	(28)	(29)	(30)	(31)	(32)
Dependent variable: Income Inequality (Atkinson index in municipalities in each province					
Urbanization rate	0.0015 *** (0.0005)	-0.0018 *** (0.0004)	-0.0014 *** (0.0004)	-0.0011 * (0.0007)	-0.0014 * (0.0008)
Square of Urbanization rate	-0.00001 ** (0.0000)	0.00002 *** (0.0000)	0.00001 *** (0.0000)	0.00001 ** (0.0000)	0.00001 ** (0.0000)
Log (GPP per capita)		-0.0215 *** (0.0026)	-0.0169 *** (0.0033)	-0.0378 *** (0.0053)	-0.0346 *** (0.0064)
*Urbanization rate * Log (GPP per capita)*					-0.0002 (0.0002)
Year FE	No	Yes	Yes	Yes	Yes
Region FE	No	No	Yes	Yes	Yes
Province (Random) Effects	No	No	No	Yes	Yes
Observations	1,386	1,377	1, 377	1, 377	1, 377
No. of provinces	77	77	77	77	77
R-Square	0.0565	0.5759	0.6111	0.3184	0.3200

Note: The time span is from 2002 to 2020; ***, **, and * indicate significance at 1%, 5%, and 10%, respectively.
Source: Author's calculation.

TABLE V
ATKINSON INDEX IN MUNICIPALITIES IN PROVINCE: CALCULATING URBAN AREAS BASED ON NIGHTTIME LIGHT DENSITY DERIVED FROM A POPULATION DENSITY OF 300 PEOPLE PER SQUARE KILOMETER

Result of Equation	(28)	(29)	(30)	(31)	(32)
Dependent variable: Income Inequality (Atkinson index in municipalities in each province					
Urbanization rate	-0.0017 *** (0.0003)	-0.0012 *** (0.0003)	-0.0005 (0.0003)	-0.0010 *** (0.0003)	-0.0011 *** (0.0003)
Square of Urbanization rate	0.00001 *** (0.0000)	0.00001 *** (0.0000)	0.00000 (0.0000)	0.00000 *** (0.0000)	0.00000 ** (0.0000)
Log (GPP per capita)		-0.0183 *** (0.0051)	-0.0173 *** (0.0052)	-0.0151 *** (0.0050)	-0.0229 *** (0.0072)
*Urbanization rate * Log (GPP per capita)*					0.0003 * (0.0001)
Year FE	No	Yes	Yes	Yes	Yes
Region FE	No	No	Yes	Yes	Yes
Province (Random) Effects	No	No	No	Yes	Yes
Observations	1,368	1,368	1,368	1,368	1,368
No. of provinces	76	76	76	76	76
R-Square	0.0365	0.8745	0.8844	0.0479	0.0489

Note: The time span is from 2002 to 2020; ***, **, and * indicate significance at 1%, 5%, and 10%, respectively.
Source: Author's calculation.

TABLE VI
ATKINSON INDEX IN MUNICIPALITIES IN PROVINCE: CALCULATING URBAN AREAS BASED ON NIGHTTIME LIGHT DENSITY DERIVED FROM A POPULATION DENSITY OF 1,500 PEOPLE PER SQUARE KILOMETER

Result of Equation	(28)	(29)	(30)	(31)	(32)
Dependent variable: Income Inequality (Atkinson index in municipalities in each province					
Urbanization rate	-0.0029 *** (0.0004)	-0.0021 *** (0.0005)	-0.0014 *** (0.0005)	-0.0015 *** (0.0005)	-0.0020 *** (0.0005)
Square of Urbanization rate	0.00003 *** (0.0000)	0.00003 *** (0.0000)	0.00002 *** (0.0000)	0.00002 *** (0.0000)	0.00002 *** (0.0000)
Log (GPP per capita)		-0.0195 *** (0.0055)	-0.0159 *** (0.0052)	-0.0160 *** (0.0052)	-0.0199 *** (0.0056)
*Urbanization rate * Log (GPP per capita)*					0.0005 *** (0.0002)
Year FE	No	Yes	Yes	Yes	Yes
Region FE	No	No	Yes	Yes	Yes
Province (Random) Effects	No	No	No	Yes	Yes
Observations	1,368	1,368	1,368	1,368	1,368
No. of provinces	76	76	76	76	76
R-Square	0.0320	0.8764	0.8888	0.0485	0.0497

Note: The time span is from 2002 to 2020; ***, **, and * indicate significance at 1%, 5%, and 10%, respectively.
Source: Author's calculation.

VI. CONCLUSION

This study has two research objectives. First, it is aimed to aggregate province-level data using conventional statistics from Thailand's NSO as the primary data sources, covering the period from 2002 to 2020. Additionally, alternative geospatial data from the U.S. Air Force DMSP and the VIIRS-DNB satellite were integrated and employed in this study. This study has two distinct definitions of urban areas: 1) Urban areas as defined by Thailand's NSO (LFS), and 2) Urban areas approximated through nighttime light data acquired from the

DMSP and VIIRS-DNB satellite. The second case of urban area determination involves two sub-categorizations: 2.1) Calculating urban areas based on nighttime light density derived from a population density of 300 people per square kilometer, and 2.2) Calculating urban areas based on nighttime light density derived from a population density of 1,500 people per square kilometer. Second, the study aims to quantitatively investigate the relationship between urbanization and income inequality in Thailand. The obtained results show that negative relationship exists between urbanization rate and income inequality in municipalities in each province measured by the Gini coefficient and Atkinson index. Moreover, the results show a negative relationship between log income and income inequality in municipalities in each province measured by the Gini coefficient and Atkinson index. On the other hand, the square of urbanization rate shows a statistically significant positive impact on income inequality in municipalities measured by the Gini coefficient and Atkinson index. In other words, urbanization rates can help reduce inequality. Especially, part of the analysis uses population density data to determine urban areas. Population density has been found to help concentrate the economy and thus reduce income inequality.

APPENDIX

A. Pool OLS

If the individual effect does not exist, OLS produce efficient and consistent parameter estimates. Assumptions of OLS are as follows:

OLS Assumptions 1. Linear in Parameters: The dependent variable (y) is a linear function of the independent variables (x) and the error term (u_i). The linear regression concerns the parameters, and it is not necessarily linear in the independent variable.

OLS Assumptions 2. Random Sampling: 1) The sample for the linear regression model must be drawn randomly from the population; 2) The number of observations in the sample should be greater than the number of parameters to be estimated in the linear regression model. This makes sense mathematically if the number of observations in the sample is less than the number of parameters, then estimation is not possible; 3) The independent variable (x) should be fixed which means independent variables should impact dependent variables. It should not be the case that dependent variables impact independent variables because regression models aim to study causal relationships rather than mere correlations between the two variables; and, 4) The error terms are random. This makes the dependent variable random.

OLS Assumptions 3. No Perfect Collinearity: In the sample, none of the independent variables (x) is constant, and there are no exact linear relationships among the independent variables (x). And, Assumption 3 allows for the independent variable (x) can be correlated but not perfectly correlated.

OLS Assumptions 4. Zero Condition Mean: The error u has the expected value of zero given any value of the explanatory variable $(E(u|x) = 0)$ or $(E(u) = 0)$. In other words, the distribution of error terms (u) has zero mean and does not depend on independent variable (x). Thus, there must be no relationship between the independent variables (x) and error terms (u).

OLS Assumptions 5. Homoskedasticity: The error terms (u) have the same variance given the other value of the explanatory variables $(Var(u|x_1, x_2, x_3, \dots, x_k) = \sigma^2)$. And, they are not related to one another (non-autocorrelation). That means the error terms (u) of different observations should not be correlated with each other $(Cov(u_i u_j|x) = 0\ for\ i \neq j)$.

If the individual effect is not equal to zero, heterogeneity may influence assumption 4 and assumption 5. The violation of assumption 4 causes random effect estimators to be biased. Thus, the OLS estimator is not the best unbiased-linear estimator. Then panel data models provide a way to deal with these problems by using linear fixed effect models and random effect models as they are more suitable.

REFERENCES

[1] Ritchie, H., & Roser, M. (2018). Urbanization. Our world in data.
[2] Wan, G., Zhang, X., & Zhao, M. (2022). Urbanization can help reduce income inequality. npj Urban Sustainability, 2(1), 1-8.
[3] Paweenawat, S. W., & McNown, R. (2014). The determinants of income inequality in Thailand: A synthetic cohort analysis. Journal of Asian Economics, 31, 10-21.
[4] Chen, G., Glasmeier, A. K., Zhang, M., & Shao, Y. (2016). Urbanization and income inequality in post-reform China: a causal analysis based on time series data. PloS one, 11(7), e0158826.
[5] Sulemana, I., Nketiah-Amponsah, E., Codjoe, E. A., & Andoh, J. A. N. (2019). Urbanization and income inequality in Sub-Saharan Africa. Sustainable Cities and Society, 48, 101544.
[6] Shutters, S. T., Applegate, J. M., Wentz, E., & Batty, M. (2022). Urbanization favors high wage earners. npj Urban Sustainability, 2(1), 1-5.
[7] Kanbur, R., & Zhuang, J. (2013). Urbanization and inequality in Asia. Asian Development Review, 30(1), 131-147.
[8] Yuan, Y., Wang, M., Zhu, Y., Huang, X., & Xiong, X. (2020). Urbanization's effects on the urban-rural income gap in China: A meta-regression analysis. Land Use Policy, 99, 104995.
[9] Ha, N. M., Dang Le, N., & Trung-Kien, P. (2021). The impact of urbanization on poverty reduction: An evidence from Vietnam. Cogent Economics & Finance, 9(1), 1918838.
[10] Wu, D., & Rao, P. (2017). Urbanization and income inequality in China: An empirical investigation at provincial level. Social Indicators Research, 131(1), 189-214.
[11] Wooldridge, J. M. (2015). *Introductory econometrics: A modern approach*. Cengage learning.