Article

The Impact of Urbanization on Income Inequality: A Study in Vietnam

Nguyen Minh Ha [1],*, Nguyen Dang Le [2] and Pham Trung-Kien [3]

[1]   Business and Economics Research Group, Ho Chi Minh City Open University, 97 Vo Van Tan Street, District 3, Ho Chi Minh City 700000, Vietnam
[2]   Department of Financial Informatics and Statistics, Ministry of Finance, Ho Chi Minh City 700000, Vietnam
[3]   Faculty of Economics, Binh Duong University, 504 Binh Duong Avenue, Binh Duong Province 590000, Vietnam
*   Correspondence: ha.nm@ou.edu.vn

check for updates

Abstract: This paper explores the impact of urbanization on income inequality in Vietnam, using the regression estimation method with panel data including Driscoll and Kraay, and Pooled Mean Group. The research data cover 63 provinces in Vietnam from 2006 to 2016. The results show that in the long term, urbanization has an impact on reducing income inequality. In the short term, urbanization has a negligible impact on income inequality. The hypothesis of an inverted-U-shaped relationship between urbanization and income inequality is confirmed. The high school enrollment rate and the proportion of agriculture have an effect on reducing income inequality.

Keywords: urbanization; income inequality; Driscoll and Kraay; PMG

1. Introduction

Many economists argue that increasing inequality is the root cause of economic crises (Stiglitz 2009). Meanwhile, the world faces serious income inequality. The gap between the rich and other groups has increased significantly. In 2015, the richest 1% of the population owned more assets than the other 99%. In Asia, the income of the poorest 70% has decreased, while that of the richest 10% has increased significantly (Hardoon et al. 2016).

The relationship between income inequality and development has long been a topic of particular interest to researchers in developed as well as developing countries. Many studies have tried to determine whether countries must make a trade-off between income inequality and growth. If so, what is the specific model of the relationship and why?

Kuznets (1955) was the first to introduce the idea of a link between inequality and development, pointing out that development involves a shift in population from traditional activities to modern activities. "An invariable accompaniment of growth in developed countries is the shift away from agriculture, a process usually referred to as industrialization and urbanization", he wrote (Kuznets 1955, p. 7). Therefore, in a simple model, income distribution among the entire population can be viewed as a combination of income distribution among those in rural and urban areas. We observe that income per capita is often lower in rural areas than in urban areas, and inequality in income distribution is lower than in rural areas in urban areas. What conclusions can we draw from these observations? First, under the same conditions, increasing the share of the urban population does not necessarily reduce economic growth: in fact, some evidence indicates that growth may be higher because urban per capita productivity increased faster in agriculture. If this is true, inequality in income distribution increases. This ideal was highlighted and clarified by Bourguignon and Morrisson (1998), Piketty (2006).

Industrialization and urbanization are closely related to economic development. The process of industrialization and urbanization affects income distribution, causing income inequality. In the short term, urbanization can increase income inequality because wages are higher for urban jobs than rural work. However, in the long term, when urbanization is highly developed, the difference in income distribution in the two regions may decrease, and income inequality will decrease.

In Vietnam today, urbanization and industrialization have progressed considerably. Vietnam's urbanization level is still low compared to the global average but, in recent years, higher than the average in other developing countries as well as countries in Southeast Asia. In terms of Gini data, income inequality in Vietnam increased until 2008 and then decreased, and in 2014 began to increase again. So, does urbanization increase income inequality? This is the question that we address in this study.

2. Literature Review

Income inequality refers to an unequal distribution of income among individuals or households. To calculate the degree of income inequality, scholars often rely on the percentage of income held by different shares of the population. Income inequality is often associated with unfairness, such as when rich people hold a significantly larger share of national income relative to their proportion of the population (Todaro 1989).

Income inequality can be measured by various methods, such as using quintiles to measure income gaps between the poorest and the richest; using World Bank Standard 40, measured with the Lorenz curve (Lorenz 1905), the Gini coefficient (Gini 1913, 1921), and the Theil index (Akita et al. 1999). In this research, we use the Gini coefficient to represent income inequality.

Urbanization concerns the physical, human, and economic development of cities. This term also includes the concentration of people and social activities on the settlement model characterized by the development of land with high population density. The result of urbanization is partly due to the increase in population, natural and by migration, as well as economic, social, and technological changes that motivate people to migrate to urban areas, which have many jobs and opportunities. Market rules and government policies promote urbanization and create related changes in people's livelihoods, land use, health, and natural resource management. Job placement decisions, rural-urban transformation and production systems, and government development and distribution policies often create urban immigration and focus on economic activities in cities (Gotham 2012).

According to Bloom et al. (2010), a basic concept commonly used to determine the level of urbanization is the proportion of the population living in urban areas, represented by Urbant and defined using the following equation:

$$Urban_t = \frac{PU_t}{PU_t + PR_t}$$

in which PU_t and PR_t represent urban and rural population respectively.

2.1. How Does Urbanization Affect Income Inequality?

An economic model including urban and rural areas has four main elements of income inequality: (1) the level of urbanization; (2) the urban-rural income gap; (3) urban income inequality; and (4) rural income inequality. By keeping urban and rural disparities and rural and urban inequalities constant, Kuznets (1955) outlines an inverted-U-shaped relationship between inequality and urbanization. In this study, urbanization is viewed as an important factor affecting income inequality, and we examine Kuznets's inverted-U-shape hypothesis based on Vietnam's provinces.

A great deal of evidence supports the idea that urbanization promotes economic growth, at least in the early stages of development, implying that a balance exists between economic growth and equal income distribution, at least geographically. Brülhart and Sbergami (2009) argue that poor countries face an awkward choice between inequality reduction and higher economic growth. In fact,

the relationship between development and income inequality described by Kuznets is highly relevant to urbanization.

The classic dual economic model examining structural change shows that inequality is an inevitable result of urbanization that is characteristic of economic development (Harris and Todaro 1970; Lewis 1954; Rauch 1993). Similarly, the New Economic Geography helps explain how economic development is associated with increased urbanization and inequality in its early stages (Krugman 1991). Both models show an increasing profit from industrial activities. In fact, many good workers are concentrated in urban areas with higher industrial wages. Economic growth is facilitated by structural changes in the economy, allowing it to enjoy the benefits of increasing profits and the economics of urbanization.

The process of urbanization brings about changes in economic structure, with people and resources being reallocated from agricultural activities to industrial activities. This process is associated with increased inequality, with higher incomes in urban areas than in rural areas. In this sense, both higher inequality and greater urbanization can enhance the concentration of production factors necessary for growth, at least in the early stages of development. And this focus further strengthens the reallocation of labor from rural to urban areas (Ross 2000). Therefore, both inequality and geographic concentration indicate, to some extent, capital accumulation (both physical and human). However, in the later stages of development, especially urban growth and the growth of a large concentration—urban accumulation—is linked to increasing inequality (Behrens and Robert-Nicoud 2014).

2.2. Previous Related Studies

Kanbur and Zhuang (2013) studied the impact of the change of two sectional economic structure and urbanization that affect inequality in Asia. Based on the Kuznets model and the inequality measurement method using the Theil index, this study analyzes four specific countries: China, India, Indonesia, and the Philippines. Their results show the following.

First, the effect of urbanization and the economic structure change on income inequality depends on the specific country. Urbanization contributes about 300% to the increase in inequality in the Philippines, more than 50% in Indonesia, and nearly 15% in India. However, it reduces inequality in China. Meanwhile, the change in the urban-rural income gap reduces inequality in Indonesia and the Philippines, but it increases inequality in India and China. Moreover, in China, the increase in income inequality is mainly due to the raise of 43% rural inequality. This is contrary to the popular point of view that increasing income inequality is the result of the expansion in the urban–rural income gap and the rising urban inequality.

Second, how urbanization will affect inequality in the future depends on the status urbanization of the country. If a country passes its "turning point", urbanization will reduce inequality in, for example, China. In contrast, urbanization will increase inequality in countries such as India, Indonesia, and the Philippines.

Third, there are four factors driving inequality in the Asian country including the degree of urbanization, urban–rural income gap, and urban and rural inequality. Among them, the urban-rural income gap is expected to have the largest marginal impact on inequality.

Sagala et al. (2014) studied the relationship between inequality in expenditure and urbanization in Indonesia. A panel data regression analysis was carried out to test the Kuznets inverted-U-shaped hypothesis based on a dataset of 33 provinces in 2000–2009. The results of the study supported the inverted-U-shaped hypothesis, regardless whether the Gini coefficient or Theil index is used as a measure of inequality. Inequality in expenditure is expected to peak at an urbanization rate of about 46–50%. Because the urbanization rate in Indonesia in 2010 was 50%, this indicates that inequality in spending has reached its highest value. Therefore, further urbanization will reduce inequality in expenditure, with other factors constant.

Oyvat (2016) studied the impact of agricultural structure and urbanization on income inequality. The author investigated the empirical relationship between inequality in land holding, urbanization,

and income inequality using cross-data sets. The estimated results indicated that the inequality of land holding has a significant impact on urbanization and urban income and inequality. Moreover, the analysis found that excessive urbanization increases income inequality. The results of the study showed that policy makers need to have a broader view of the importance of agricultural policies. An issue of progressive land reform and subsidies to protect small farmers rogressive land reform and subsidies to protect small farmers could also reduce urban income inequality and poverty in the long term.

Wu and Rao (2017) studied inequality in China, focusing on identifying the main causes of inequality. The main objective of the study was to examine the relationship between urbanization and income inequality using provincial data. Panel data in 20 provinces was collected from the China Statistical Yearbook for five years include 1998, 2000, 2002, 2005, and 2010. The empirical analysis was based on ordinary least squares estimator and fixed and random effects models, showing a strong inverted-U-shaped relationship between inequality and urbanization. An urbanization rate of 0.53 has been determined, with the implication that provinces with higher levels of urbanization can reduce income inequality.

However, Angeles (2010) used urban population density to represent the urbanization rate, and its square as an explanatory variable in the regression analysis of panel data on income inequality. With panel data on 226 countries and regions in 1960–2005, a U-shaped relationship was found, not an inverted-U-shaped relationship. Although it is not statistically significant, this result does not support Kuznets's hypothesis.

3. Methodology and Data

3.1. Empirical Framework

Based on the theory related to urbanization and income inequality, the impact of urbanization on income equality, and the research model of Sagala et al. (2014), Wu and Rao (2017), and related studies, we present a quantitative model assessing the impact of urbanization on income inequality as follows:

$$INEQ_{it} = \alpha + \beta_1 URB_{it} + \beta_2 URB\_sq_{it} + \beta_3 GRDPpc_{it} + \beta_4 GRDPpc\_sq_{it} +$$
$$\beta_5 initialINEQGRDPpc_{it} + \beta_6 rEXP_{it} + \beta_7 rGOV_{it} + \beta_8 rEDU_{it} + \beta_9 rAGR_{it} + \varepsilon_{it}$$

(1)

Definition of variables in research model are summarized in Table 1.

Table 1. Definition of variables in research model.

Variable Label	Definition	Expected Sign
	Dependent variable	
INEQ	Gini index	
	Independent variable	
URB	Urban population as a share of the average population in the province	+
URB_sq	Square of *URB*	−
GRDPpc	Gross regional domestic product per capita	+
GRDPpc_sq	Square of *GRDPpc*	−
initialINEQGRDPpc	Variable interaction between initial inequality and *GRDPpc*	
rEXP	Export value as a share of the province's *GRDP*	−
rGOV	Public expenditure as a share of the province's *GRDP*	−
rEDU	Students entering high school as a share of the average population in the province	−
rAGR	Agricultural value as a share of the province's *GRDP*	−

3.2. Data

Due to limited access to Vietnam's statistical data and availability of the dataset Vietnam Household Living Standard Survey (VHLSS), this study uses panel data on 63 provinces in Vietnam for every two years from 2006 to 2016. The total number of observations is 378 (6 years × 63 provinces).

The data used in our quantitative analysis include the Gini coefficient, representing income inequality, which was calculated by the author from VHLSS, and secondary data was collected from reports and surveys conducted by the General Statistics Office of Vietnam every two years.

The research uses urban population as a share of total population to proxy for urbanization. This variable is commonly used in previous studies (Nguyen and Nguyen 2018).

Data for independent and control variables used in quantitative models include urbanization, gross regional domestic product (GRDP) per capita, export volume as a share of the province's GRDP, public expenditure as a share of the province's GRDP, students entering high school as a share of the average population of the province, and agricultural value as a share of the province's GRDP, which were collected from the annual statistical yearbook of the General Statistics Office and the Provincial Statistics Office in Vietnam.

The study uses regression methods to estimate static panel data with OLS, FE, RE, and regression correction techniques proposed by Driscoll and Kraay (1998) and pooled mean group (PMG) estimation (Pesaran et al. 1999; Pesaran and Smith 1995).

4. Empirical Results

4.1. Descriptive Statistics of Variables in the Research Model

Table 2 summarizes the descriptive statistics of all the variables used in the model. The mean value of *INEQ* is 0.3794, its standard deviation is 0.0537, its minimum is 0.2498, and its maximum is 0.5883. Thus, our sample does not have much difference in the income inequality index. For *URB*, the mean is 0.2597, the standard deviation is 0.1640, the minimum is 0.0736, and the maximum is 0.8746—this shows a significant disparity in the level of urbanization among provinces during the study period. Like *URB*, *GRDPpc*, *rEXP*, *rGOV*, and *rAGR* also have large differences among provinces. For *rEDU*, the average is 0.0313, and the standard deviation is 0.0082, indicating a relatively small difference.

Table 2. Descriptive Statistics.

Variable	Obs.	Mean	S.D.	Min.	Max.
INEQ	378	0.3794	0.0537	0.2498	0.5883
URB	378	0.2597	0.1640	0.0736	0.8746
GRDPpc	378	0.3002	0.3693	0.0356	3.9169
rEXP	378	0.4413	0.6613	0.7863	6.2757
rGOV	378	0.3450	0.2512	0.1322	1.8091
rEDU	378	0.0313	0.0082	0.0123	0.0575
rAGR	378	0.2870	0.1444	0.0083	0.6241

4.2. Correlation Matrix and Multicollinearity

The correlation between variables in the regression model indicates the presence of multicollinearity that can affect the accuracy of the regression results. The results of the correlation analysis between variables in Table 3 show that the pairs of independent variables are not significantly correlated. However, to ensure accurate estimation results, we conduct a test on the multicollinearity phenomenon between variables in Table 4.

Table 3. Correlation matrix.

Variable	URB	GRDPpc	rEXP	rGOV	rEDU	rAGR
URB	1.000					
GRDPpc	0.4223 *	1.000				
rEXP	0.2164 *	0.2254 *	1.000			
rGOV	−0.2579 *	−0.2675 *	−0.3093 *	1.000		
rEDU	−0.100	−0.1489 *	−0.1449 *	−0.050	1.000	
rAGR	−0.5279 *	−0.4605 *	−0.3387 *	0.1350 *	−0.101	1.000

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

In Table 4, the results show that the coefficient VIF of the variables is quite small (<2), so the model has no multicollinearity.

Table 4. Checking multicollinearity.

Variable	VIF	SQRT VIF	Tolerance	R-Squared
URB	1.56	1.25	0.6417	0.3583
GRDPpc	1.46	1.21	0.6855	0.3145
rEXP	1.29	1.14	0.7753	0.2247
rGOV	1.24	1.11	0.8074	0.1926
rEDU	1.15	1.07	0.8678	0.1322
rAGR	1.80	1.34	0.555	0.445
Mean VIF	1.42			

4.3. Assessing the Impact of Urbanization on Income Inequality

First, the implementation of static panel data regression estimation methods follows that in model (1). To assess the impact of urbanization (*URB*) on income inequality (*INEQ*), without considering the relevant factors, we revised this model by excluding *URB_sq*, *GRDPpc_sq*, and *initialINEQGRDPpc* in model (2):

$$INEQ_{it} = \alpha + \beta_1 URB_{it} + \beta_3 GRDPpc_{it} + \beta_6 rEXP_{it} + \beta_7 rGOV_{it} + \beta_8 rEDU_{it} + \beta_9 rAGR_{it} + \varepsilon_{it} \quad (2)$$

To test the inverted-U-shaped hypothesis between income inequality and urbanization, we added *URB_sq* to model (2), revised as model (3):

$$INEQ_{it} = \alpha + \beta_1 URB_{it} + \beta_2 URB\_sq_{it} + \beta_3 GRDPpc_{it} + \beta_6 rEXP_{it} + \beta_7 rGOV_{it} + \beta_8 rEDU_{it} + \beta_9 rAGR_{it} + \varepsilon_{it} \quad (3)$$

To test the inverted-U-shaped hypothesis between income inequality and economic growth, we added *GRDPpc_sq* to model (2), revised as model (4):

$$INEQ_{it} = \alpha + \beta_1 URB_{it} + \beta_3 GRDPpc_{it} + \beta_4 GRDPpc\_sq_{it} + \beta_6 rEXP_{it} + \beta_7 rGOV_{it} + \beta_8 rEDU_{it} + \beta_9 rAGR_{it} + \varepsilon_{it} \quad (4)$$

To analyze the effects of initial inequality on the impact of economic growth on income inequality, we added *initialINEQGRDPpc* to model (2), revised as model (5):

$$INEQ_{it} = \alpha + \beta_1 URB_{it} + \beta_3 GRDPpc_{it} + \beta_5 initialINEQGRDPpc_{it} + \beta_6 rEXP_{it} + \beta_7 rGOV_{it} + \beta_8 rEDU_{it} + \beta_9 rAGR_{it} + \varepsilon_{it} \quad (5)$$

Finally, to analyze the impact of *URB* on *INEQ* under the simultaneous influence of all factors, we constructed model (6):

$$INEQ_{it} = \alpha + \beta_1 URB_{it} + \beta_2 URB\_sq_{it} + \beta_3 GRDPpc_{it} + \beta_4 GRDPpc\_sq_{it} + \beta_5 initialINEQGRDPpc_{it} + \beta_6 rEXP_{it} + \beta_7 rGOV_{it} + \beta_8 rEDU_{it} + \beta_9 rAGR_{it} + \varepsilon_{it} \quad (6)$$

To select an estimation method that is consistent with the dataset, we chose between pooled OLS and FE methods through F-testing with the hypothesis "H_0: Pooled OLS method is appropriate ($u_i = 0$)". The regression result of the FE method in Table 5 shows that "F-test $u_i = 0$" has a *p*-value = 0.000 < 0.01 in all models, which indicates that the FE method is more appropriate than pooled OLS. The results in Table 5 also show that the F-test has a *p*-value < 0.01 in all models, indicating that at the 1% significance level, the models are statistically significant. Then, to choose between FE and RE, the study uses the Hausman test with the hypothesis "H_0: RE model is appropriate". The test results show (Prob > Chi2) = 0.0000 < 0.01 in all models, which demonstrates sufficient statistical evidence to reject H_0, indicating that the more efficient FE model should be selected.

However, the results of post-regression tests with the FE model have the following violations: aWald test with the hypothesis "H_0: constant variance" gives the result Prob > Chi2 = 0.0000 < 0.01 in all empirical models, rejecting H_0 at a significance level of 1%, which means that the existing models of variance change. The Wooldridge autocorrelation test with the hypothesis "H_0: no autocorrelation" gives the statistical result Prob > F = 0.0000 < 0.05 in all models, rejecting H_0 at a significance level of 5%, which means that the model has autocorrelation. The result from testing Pesaran cross-sectional dependence with the hypothesis "H_0: Cross-sectional independence" is (Prob > z) <0.01 in all models, rejecting the H_0 hypothesis at a significance level of 1%, which means that cross-sectional dependence exists in models. The study also uses the Durbin-Wu-Hausman test for endogeneity of explanatory variables with the hypothesis "H_0: explanatory variables are exogenous," i.e., not correlated with residuals. The results show that *rEXP* and *rGOV* are endogenous in the model.[1]

Table 5. Results of fixed effects regression.

Variables	(2)	(3)	(4)	(5)	(6)
	INEQ	*INEQ*	*INEQ*	*INEQ*	*INEQ*
URB	−0.307 ***	−0.00200	−0.247 ***	−0.323 ***	0.0657
	(0.0680)	(0.237)	(0.0727)	(0.0728)	(0.238)
URB_sq		−0.319			−0.353
		(0.237)			(0.234)
GRDPpc	−0.0233	−0.0229	−0.0891 ***	0.0631	0.289 *
	(0.0151)	(0.0150)	(0.0327)	(0.144)	(0.158)
GRDPpc_sq			0.0150 **		0.0281 ***
			(0.00662)		(0.00824)
initialINEQGRDPpc				−0.215	−1.083 **
				(0.356)	(0.439)
rEXP	−0.00584	−0.00753	−0.00527	−0.00619	−0.00843
	(0.00504)	(0.00518)	(0.00501)	(0.00507)	(0.00513)
rGOV	0.00847	0.00820	0.00416	0.00934	0.00440
	(0.0202)	(0.0202)	(0.0202)	(0.0203)	(0.0200)
rEDU	−1.302 **	−1.006	−2.133 ***	−1.182 *	−1.932 **
	(0.658)	(0.692)	(0.750)	(0.687)	(0.766)
rAGR	−0.0839 *	−0.0746	−0.124 **	−0.0821 *	−0.140 ***
	(0.0468)	(0.0472)	(0.0497)	(0.0469)	(0.0505)
Constant	0.531 ***	0.470 ***	0.570 ***	0.528 ***	0.527 ***
	(0.0339)	(0.0563)	(0.0380)	(0.0342)	(0.0578)
Observations	378	378	378	378	378
F-test (*p*-value)	0.0000	0.0000	0.0000	0.0001	0.0000
R-squared	0.093	0.098	0.108	0.094	0.132
F-test $u_i = 0$ (*p*-value)	0.0000	0.0000	0.0000	0.0000	0.0000
Number of id	63	63	63	63	63

Standard errors in parentheses: *** *p* < 0.01, ** *p* < 0.05, * *p* < 0.1.

---

[1] The results of estimating OLS and RE regression, Hausman test, and post-regression tests are detailed in the Appendix A.

The models have an endogenous element caused by *rEXP* and *rGOV*, but this is not the main research focus, so we ignore this problem.

A regression model can be used for analysis of the estimation results, and statistical inference techniques should be used for appropriate revisions. Technical studies revise regressions following Driscoll and Kraay (1998), whose method resolves violations in the FE model, such as changes in the variance, autocorrelation, and cross-dependence. The results are in Table 6.

Finally, the impact of urbanization on income inequality in the short and long term is assessed. The technical econometric dynamic panel regression estimatation method used in this study is PMG (Pesaran et al. 1999; Pesaran and Smith 1995). Because the number of years of observation is not long enough, the study assesses only the impact of urbanization on income inequality unaffected by other factors. The results in Table 7 show that urbanization affects income inequality in the long term but not in the short term.

Table 6. Results of Driscoll and Kraay regression.

Variables	(2)	(3)	(4)	(5)	(6)
	INEQ	*INEQ*	*INEQ*	*INEQ*	*INEQ*
URB	−0.307 ***	−0.00200	−0.247 **	−0.323 ***	0.0657
	(0.0729)	(0.0489)	(0.0907)	(0.0708)	(0.0419)
URB_sq		−0.319 **			−0.353 **
		(0.0920)			(0.103)
GRDPpc	−0.0233 **	−0.0229 **	−0.0891 *	0.0631	0.289
	(0.00884)	(0.00882)	(0.0370)	(0.194)	(0.194)
GRDPpc_sq			0.0150		0.0281 ***
			(0.00771)		(0.00675)
initialINEQGRDPpc				−0.215	−1.083 *
				(0.484)	(0.484)
rEXP	−0.00584	−0.00753	−0.00527	−0.00619	−0.00843
	(0.00579)	(0.00615)	(0.00619)	(0.00638)	(0.00693)
rGOV	0.00847	0.00820	0.00416	0.00934	0.00440
	(0.0121)	(0.0127)	(0.00794)	(0.0107)	(0.00804)
rEDU	−1.302	−1.006	−2.133 ***	−1.182 *	−1.932 ***
	(0.722)	(0.696)	(0.302)	(0.469)	(0.338)
rAGR	−0.0839 ***	−0.0746 ***	−0.124 **	−0.0821 **	−0.140 **
	(0.0197)	(0.0166)	(0.0459)	(0.0218)	(0.0437)
Constant	0.531 ***	0.470 ***	0.570 ***	0.528 ***	0.527 ***
	(0.0297)	(0.0219)	(0.0111)	(0.0266)	(0.0160)
Observations	378	378	378	378	378
Number of groups	63	63	63	63	63

Standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 7. Results of PMG regression.

Variables	Long Run	Short Run
__ec		−0.889 ***
		(0.0959)
D.URB		−1.199
		(1.940)
URB	−0.500 ***	
	(0.0129)	
Constant		0.457 ***
		(0.0466)
Observations	315	315

Standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

5. Discussion of Results

The results of the econometric models lead to the following findings.

URB has a negative regression coefficient with a statistical significance of 1% in model (2), indicating the negative impact of urbanization on income inequality in Vietnamese provinces. This means that higher urbanization contributes to reducing income inequality, which is consistent with the impact of urbanization on income inequality in models (3), (4), and (5). This result is consistent with Johansson and Wang (2014) and contrary to Beladi et al. (2017). In fact, in Vietnam, urbanization is associated with the formation of industrial zones and clusters. People who have little or no land in rural areas migrate to cities to work in factories with higher wages than previous jobs in rural areas, which raises their income. Therefore, urbanization contributes to reducing income inequality in Vietnam. This impact is consistent with the PMG method in the long term.

URB_sq, the impact of urbanization on income inequality, may be nonlinear, in other words, this effect may be in accordance with the inverted-U-shaped hypothesis. *URB_sq* is included in models (3) and (6) to test this hypothesis, and the estimation results show that its impact on *INEQ* is negatively significant at the 5% level. Model (6), which includes with regression coefficients for *URB* and *URB_sq*, shows the nonlinear relationship between urbanization and income inequality. Urbanization increases income inequality in the early stages until it reaches a certain threshold, after which it reduces income inequality. This result is consistent with Sagala et al. (2014) in Indonesia and Wu and Rao (2017) in China.

GRDPpc, which has a negative regression coefficient with a statistical significance of 5% in model (2), shows the negative impact of per capita income on income inequality in Vietnam provinces, which means that higher per capita income contributes to reducing income inequality. This result is consistent with the impact of per capita income on income inequality in models (3) and (4), consistent with Jin (2009) who studies China in 1990–2006. In addition, the interaction variable between initial economic growth and inequality (*initialINEQGRDPpc*) has no significant regression coefficients. The results of model (5) show that the effect of economic growth on income inequality in Vietnam in 2006–2016 is not affected by the level of initial inequality. The results are contrary to those of Johansson and Wang (2014).

The regression coefficients of *GRDPpc_sq*, which are not statistically significant in model (4), shows that the inverted-U-shape hypothesis between economic growth and income inequality in Vietnam is not confirmed from 2006 to 2016. The results are contrary to those of Sagala et al. (2014), who found evidence supporting the inverted-U-shape hypothesis in Indonesia between 2000 and 2009.

rEXP, which is not statistically significant, shows that trade, which represents the share of exports in GDP, has a negligible effect on income inequality in Vietnam. This result is similar to that of Johansson and Wang (2014).

rGOV, which is not statistically significant, shows that the government spending in the provinces has a negligible effect on income inequality in Vietnam. This result is similar to that of Johansson and Wang (2014) and Beladi et al. (2017).

rEDU, which represents human capital, has a negative regression coefficient, showing the negative impact of education on income inequality. Although the level of statistical significance is not consistent in all empirical models, the results show that education plays a role in decrease income inequality in Vietnam. This result contrasts with Jin (2009), showing the education increased income inequality in China between 1990 and 2006, with Johansson and Wang (2014), which found no impact of education on income inequality in 90 countries in 1981–2005.

rAGR, which has a negative regression coefficient and is statistically significant at the 1% significance level in models (2) and (3) and 5% in models (4) and (5) shows the negative effect of the proportion of agricultural value to income inequality. This result is consistent with that of Wu and Rao (2017).

6. Conclusions and Policy Implications

This study analyzed and evaluated the impact of urbanization on income inequality in Vietnam in the period 2006–2016. We used the following econometric techniques and methods: the estimation methods with static panel data regression used are OLS, FE, RE, and Driscoll and Kraay. The method used for estimating dynamic panel data regression is PMG, enabling us to consider the impact of urbanization on income inequality in the short and long term.

Our research results lead us to draw the following conclusions. Urbanization helps reduce income inequality in Vietnam, including in the long term. We confirm the inverted-U-shaped relationship between urbanization and income inequality in Vietnam, where economic growth reduces income inequality. The impact of economic growth on income inequality is not affected by the initial level of inequality. The hypothesis on an inverted-U-shaped relationship between economic growth and income inequality is not confirmed during the study period. The share of exports has a negligible effect on income inequality in Vietnam. Public spending does not significantly affect income inequality in Vietnam. The high school enrollment rate and the proportion of agriculture influence reductions in income inequality.

Because urbanization has the effect of reducing income inequality, even in the long term, Vietnam needs to continue to promote urbanization. Economic growth has the effect of reducing income inequality, therefore policies to increase economic growth and social welfare should be maintained to reduce the gap between rich and poor. As improving intellectual standards will help reduce inequality, more supportive policies are needed to improve education. Because high value agricultural development will contribute to reducing inequality, appropriate policies are needed to support and develop agriculture, farmers, and rural areas.

Appendix A

Table A1. Results of pooled OLS regression.

Variables	(2) INEQ	(3) INEQ	(4) INEQ	(5) INEQ	(6) INEQ
URB	−0.00142 (0.0191)	0.0299 (0.0585)	0.00526 (0.0194)	0.00115 (0.0184)	−0.00203 (0.0573)
URB_sq		−0.0371 (0.0654)			0.00174 (0.0637)
GRDPpc	0.0146* (0.00822)	0.0140 * (0.00830)	−0.0219 (0.0222)	−0.415 *** (0.0787)	−0.421 *** (0.0804)
GRDPpc_sq			0.0106 * (0.00598)		−0.00286 (0.00636)
initialINEQGRDPpc				1.036 *** (0.189)	1.074 *** (0.209)
rEXP	−0.00500 (0.00431)	−0.00521 (0.00433)	−0.00406 (0.00434)	0.00191 (0.00434)	0.00193 (0.00438)
rGOV	0.0799 *** (0.0111)	0.0803 *** (0.0112)	0.0763 *** (0.0113)	0.0699 *** (0.0109)	0.0705 *** (0.0110)
rEDU	−0.642 * (0.328)	−0.647 ** (0.328)	−0.854 ** (0.348)	−1.021 *** (0.323)	−0.977 *** (0.338)
rAGR	0.0507 ** (0.0234)	0.0493 ** (0.0235)	0.0351 (0.0249)	0.0246 (0.0230)	0.0279 (0.0243)
Constant	0.356 *** (0.0191)	0.352 *** (0.0204)	0.374 *** (0.0218)	0.388 *** (0.0193)	0.384 *** (0.0222)
Observations	378	378	378	378	378
R-squared	0.189	0.189	0.195	0.250	0.250

Standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A2. Results of RE regression.

Variables	(2) INEQ	(3) INEQ	(4) INEQ	(5) INEQ	(6) INEQ
URB	−0.0314 (0.0258)	0.0527 (0.0812)	−0.0217 (0.0262)	−0.0193 (0.0239)	0.0273 (0.0745)
URB_sq		−0.0990 (0.0896)			−0.0514 (0.0822)
GRDPpc	−0.00198 (0.0100)	−0.00349 (0.0101)	−0.0510 ** (0.0230)	−0.388 *** (0.0903)	−0.376 *** (0.0931)
GRDPpc_sq			0.0135 ** (0.00574)		0.00272 (0.00637)
initialINEQGRDPpc				0.949 *** (0.219)	0.895 *** (0.245)
rEXP	−0.00627 (0.00456)	−0.00682 (0.00459)	−0.00557 (0.00454)	−0.00156 (0.00455)	−0.00185 (0.00459)
rGOV	0.0564 *** (0.0136)	0.0568 *** (0.0136)	0.0520 *** (0.0136)	0.0535 *** (0.0129)	0.0544 *** (0.0129)
rEDU	−0.682 * (0.395)	−0.668 * (0.397)	−1.110 ** (0.434)	−1.125 *** (0.388)	−1.166 *** (0.408)
rAGR	0.0153 (0.0286)	0.0130 (0.0288)	−0.0126 (0.0308)	−0.00103 (0.0276)	−0.00458 (0.0290)
Constant	0.388 *** (0.0226)	0.377 *** (0.0253)	0.420 *** (0.0261)	0.413 *** (0.0226)	0.409 *** (0.0266)
Observations	378	378	378	378	378
Number of id	63	63	63	63	63

Standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A3. Hausman test results.

Statistical Parameters	(2)	(3)	(4)	(5)	(6)
Chi2	58.27	58.63	55.59	49.78	68.39
Prob > chi2	0.0000	0.0000	0.0000	0.0000	0.0000

Table A4. Wald test of constant variance.

Statistical Parameters	(2)	(3)	(4)	(5)	(6)
Chi2	8920.95	12,922.50	6332.03	9820.80	16,898.79
Prob > chi2	0.0000	0.0000	0.0000	0.0000	0.0000

Table A5. Wooldridge test inspection of autocorrelation.

Statistical Parameters	(2)	(3)	(4)	(5)	(6)
$F_{(1,62)}$	6.297	5.556	6.457	5.997	5.207
Prob>F	0.0147	0.0216	0.0136	0.0172	0.0259

Table A6. Pesaran test inspection dependence between units.

Statistical Parameters	(2)	(3)	(4)	(5)	(6)
Pesaran's test	3.915	3.520	3.945	3.979	3.675
Prob > z	0.0001	0.0004	0.0001	0.0001	0.0002
abs	0.401	0.403	0.395	0.403	0.397

Table A7. Durbin-Wu-Hausman test on the endogenousness of explanatory variables.

Variables	(2)	(3)	(4)	(5)	(6)
URB					
Chi-sq	0.36863	0.41196	0.37807	0.21610	0.19385
p-value	0.54375	0.52098	0.53864	0.64203	0.65973
URB_sq					
Chi-sq		0.61870			0.30912
p-value		0.43153			0.57822
GRDPpc					
Chi-sq	0.37675	0.38193	0.02666	0.88462	0.67078
p-value	0.53935	0.53657	0.87030	0.34694	0.41278
GRDPpc_sq					
Chi-sq			0.75805		0.05507
p-value			0.38394		0.81446
initial INEQGRDPpc					
Chi-sq				0.64876	0.46360
p-value				0.42056	0.49595
rEXP					
Chi-sq	3.67900	3.65990	4.37889	7.78901	7.89755
p-value	*0.05510*	*0.05574*	*0.03639*	*0.00526*	*0.00495*
rGOV					
Chi-sq	3.91847	3.93987	3.64679	3.52257	3.44806
p-value	*0.04776*	*0.04715*	*0.05618*	*0.06054*	*0.06333*
rEDU					
Chi-sq	0.02895	0.02656	0.12883	0.09891	0.13693
p-value	0.86490	0.87053	0.71965	0.75314	0.71135
rAGR					
Chi-sq	0.07408	0.07707	0.00048	0.03788	0.06896
p-value	0.78548	0.78131	0.98260	0.84569	0.79286

References

Akita, Takahiro, Rizal Affandi Lukman, and Yukino Yamada. 1999. Inequality in the Distribution of Household Expenditures in Indonesia: A Theil Decomposition Analysis. *The Developing Economies* 37: 197–221. [CrossRef]

Angeles, Luis. 2010. An alternative test of Kuznets' hypothesis. *The Journal of Economic Inequality* 8: 463–73. [CrossRef]

Behrens, Kristian, and Frédéric Robert-Nicoud. 2014. Survival of the fittest in cities: agglomeration, polarization, and inequality. *The Economic Journal* 124: 1371–400. [CrossRef]

Beladi, Hamid, Chi-Chur Chao, Mong Shan Ee, and Daniel Hollas. 2017. *Urban Development, Excessive Entry of Firms and Wage Inequality in Developing Countries.* ADBI Working Paper Series, ADBI Working Paper 653; Tokyo: Asian Development Bank Institute.

Bloom, David E., David Canning, Günther Fink, Tarun Khanna, and Patrick Salyer. 2010. *Urban Settlement: Data, Measures, and Trends.* UNU-WIDER Working Paper. Helsinki: World Institute for Development Economic Research (UNU-WIDER).

Bourguignon, François, and Christian Morrisson. 1998. Inequality and Development: The Role of Dualism. *Journal of Development Economics* 57: 233–57. [CrossRef]

Brülhart, Marius, and Federica Sbergami. 2009. Agglomeration and growth: cross-country evidence. *Journal of Urban Economics* 65: 48–63. [CrossRef]

Driscoll, John C., and Aart C. Kraay. 1998. Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *Review of Economics and Statistics* 80: 549–60. [CrossRef]

Gini, Corrado. 1913. Variabilita e Mutabilita. *Journal of the Royal Statistical Society* 76: 326–27.

Gini, Corrado. 1921. Measurement of inequality of incomes. *The Economic Journal* 31: 124–26. [CrossRef]

Gotham, Kevin Fox. 2012. Urbanization. In *The Wiley-Blackwell Companion to Sociology.* Edited by G. Ritzer Malden. Hoboken: Wiley-Blackwell, pp. 488–503.

Hardoon, Deborah, Sophia Ayele, and Ricardo Fuentes-Nieva. 2016. *An Economy for the 1%.* Nairobi: Oxfam.

Harris, John R., and Michael P. Todaro. 1970. Migration, Unemployment, and Development: A Two Sector Analysis. *American Economic Review* 60: 126–42.

Jin, Furong. 2009. Foreign Direct Investment and Income Inequality in China. *Seoul Journal of Economics* 22: 311–39.

Johansson, Anders C., and Xun Wang. 2014. Financial sector policies and income inequality. *China Economic Review* 31: 367–78. [CrossRef]

Kanbur, Ravi, and Juzhong Zhuang. 2013. Urbanization and inequality in Asia. *Asian Development Review* 30: 131–47. [CrossRef]

Krugman, Paul R. 1991. *Geography and Trade.* Cambridge: MIT Press.

Kuznets, Simon. 1955. Economic Growth and Income Inequality. *American Economic Review* 45: 1–28.

Lewis, W. Arthur. 1954. Economic Development with Unlimited Supplies of Labor. *Manchester School* 22: 139–91. [CrossRef]

Lorenz, Max O. 1905. Methods of Measuring the Concentration of Wealth. *Publications of the American Statistical Association* 9: 209–19. [CrossRef]

Nguyen, Ha Minh, and Le Dang Nguyen. 2018. The relationship between urbanization and economic growth: An empirical study on ASEAN countries. *International Journal of Social Economics* 45: 316–39. [CrossRef]

Oyvat, C. 2016. Agrarian Structures, Urbanization, and Inequality. *World Development* 83: 207–30. [CrossRef]

Pesaran, M. Hashem, and Ron Smith. 1995. Estimating Long-run Relationships from Dynamic Heterogeneous Panels. *Journal of Econometrics* 68: 79–113. [CrossRef]

Pesaran, M. Hashem, Yongcheol Shin, and Ron P. Smith. 1999. Pooled Mean Group Estimation of Dynamic Heterogeneous Panels. *Journal of the American Statistical Association* 94: 621–34. [CrossRef]

Piketty, Thomas. 2006. The Kuznets Curve, Yesterday and Tomorrow. In *Understanding Poverty.* Edited by Abhijit Banerjee, Roland Benabou and Dilip Mookherjee. New York: Oxford University Press, pp. 63–72.

Rauch, James E. 1993. Economic development, urban underemployment, and income inequality. *Canadian Journal of Economics* 26: 901–18. [CrossRef]

Ross, Jaime. 2000. *Development Theory and the Economics of Growth.* Ann Arbor: The University of Michigan Press.

Sagala, Perdamen, Takahiro Akita, and Arief Anshori Yusuf. 2014. Urbanization and expenditure inequality in Indonesia: Testing the Kuznets hypothesis with provincial panel data. *Letters in Spatial and Resource Sciences* 7: 133–47. [CrossRef]

Stiglitz, Joseph. 2009. The global crisis, social protection and jobs. *International Labour Review* 148: 1–2. [CrossRef]

Todaro, Michael P. 1989. *Economic development in the Third World*, 4th ed. London and New York: Longman.

Wu, Dongjie, and Prasada Rao. 2017. Urbanization and Income Inequality in China: An Empirical Investigation at Provincial Level. *Social Indicators Research* 131: 189–214. [CrossRef]