



WIDER Working Paper 2018/106

Earnings polarization, ethnicity, and regional perspective in Indonesia

Arip Muttaqien¹, Denisa Sologon², and Cathal O'Donoghue³

September 2018

Abstract: Recently, quantitative methods have been increasingly used in ethnicity research, which traditionally has relied mainly on qualitative methods. However, quantitative studies on ethnicity in Indonesia are scarce, even though the country has more than 600 ethnic groups living across some 17 thousand islands and a history of ethnic conflicts in several regions. This study aims to address the earnings polarization in Indonesia, which is interwoven with social problems, ethnic conflicts, and social tensions. In particular, we examine the impact of ethnic diversity on earnings polarization in the Indonesian labour market using Re-centered Influence Function regression approach. With considering additional covariates, the results show that regional characteristics are more important than ethnicity. Finally, ethnicity becomes not significant anymore by including the interaction effect between ethnicity and regional characteristics.

Keywords: ethnicity, earnings polarization, regional characteristics, Indonesia

JEL classification: J08, O53

Acknowledgements: We are thankful to Pia Rattenhuber for valuable comments, Philippe van Kerm for insightful technical discussion of RIF-approach, and Araar Abdelkrim for valuable technical discussion of DASP. We would like to thank Erlangga Landiyanto, Talitha Chairunissa, and Muhammad Sowwam for important comments and enjoyable discussions. We are greatly indebted for the participants' feedback in UNU-WIDER seminar (May 2018), Lab Policy of NUI Galway (May 2018), and internal conference at UNU-MERIT (June 2018). The authors are grateful to UNU-WIDER for its support in completing this working paper. Lastly, Arip Muttaqien gratefully acknowledges support from the Indonesian Endowment Fund for Education (LPDP) – Indonesian Ministry of Finance.

¹ United Nations University - Maastricht Economic and Social Research Institute on Innovation and Technology (UNU-MERIT), Maastricht, the Netherlands, corresponding author: muttaqien@merit.unu.edu; ² Luxembourg Institute of Socio-Economic Research (LISER); ³ The College of Arts, Social Sciences, and Celtic Studies, National University of Ireland – Galway (NUI Galway).

This study has been prepared within the UNU-WIDER Visiting PhD Fellowship programme.

Copyright © The Authors 2018

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9256-548-0 <https://doi.org/10.35188/UNU-WIDER/2018/548-0>

Typescript prepared by Ans Vehmaanperä.

The United Nations University World Institute for Development Economics Research provides economic analysis and policy advice with the aim of promoting sustainable and equitable development. The Institute began operations in 1985 in Helsinki, Finland, as the first research and training centre of the United Nations University. Today it is a unique blend of think tank, research institute, and UN agency—providing a range of services from policy advice to governments as well as freely available original research.

The Institute is funded through income from an endowment fund with additional contributions to its work programme from Finland, Sweden, and the United Kingdom as well as earmarked contributions for specific projects from a variety of donors.

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

The views expressed in this paper are those of the author(s), and do not necessarily reflect the views of the Institute or the United Nations University, nor the programme/project donors.

1 Introduction

Quantitative analyses of ethnic diversity are increasingly applied (Fearon 2003). For instance, ethnic diversity has been used as one of explanatory variables to examine the provision of public goods (Alesina et al. 1999; Habyarimana et al. 2007; Schündeln 2013), deforestation (Alesina et al. 2014), democratization (Merkel and Weiffen 2012), and social conflict (Esteban and Schneider 2008; Esteban and Ray 2008; Esteban et al. 2012).

Indonesia is one of the most ethnically diverse countries in Asia. Despite this rich ethnic diversity, however, few studies have examined ethnicity in Indonesia using a quantitative approach (Arifin et al. 2015). According to the full Indonesian census dataset of 243 million observations in 2010, Ananta et al. (2015) and Arifin et al. (2015) re-classified 1,331 ethnic categories into more than 600 ethnic groups. These two studies thereby constructed ethnic diversity at the district level throughout the archipelago of Indonesia. Additionally, Indonesia ranks amongst the ethnically and linguistically most heterogeneous countries in Asia (Fearon 2003).

Analysing ethnicity in Indonesia is necessary from a historical perspective as well. Before the independence of the nation in 1945, the Dutch East Indies colonial government segregated people into a higher class (i.e., European immigrants) and a lower class (i.e., indigenous/original people) (Beets et al. 2002). After independence until 1990s, public discussion of ethnicity was a political taboo and considered as politically incorrect in Indonesia, especially under the New Order regime (1966–1988) (Ananta et al. 2014; Arifin et al. 2017). Recent policy has aimed to strengthen social and political stability in an effort to unify the ethnically diverse nation and thereby overcome the ethnic discrimination experienced in the colonial era.

Indonesia has also experienced violent ethnic conflict in the past. Especially during the 1990s, the country faced ethnic violence that almost destroyed the national unity (Bertrand 2004). As Bertrand (2004) explained, the riots and high tensions across ethnic groups in some regions caused numerous deaths and destruction of private assets. At least 10,000 people were killed from 1997 to 2002. For instance, the violence towards ethnically Chinese groups surged during that period. The violence between the indigenous ethnic group of the Dayak and the immigrant ethnic group of the Maduranese resulted in at least 1,000 deaths in Kalimantan. The conflict between Christians and Muslims escalated and resulted in at least 5,000 deaths from 1999 to 2002 in the Maluku region. Other conflicts occurred in Aceh, which is located on the edge of the western region; in the independent Timor Leste, which is a former province of Indonesia; and in Papua, which is on the edge of the eastern region.

The previous literature on ethnicity in Indonesia has mostly focused on ethnicity itself without considering economic variables in the analysis, such as in Ananta et al. (2014), Ananta et al. (2015), and Arifin et al. (2015). The latest contribution, provided by Muller (2016), analyzed ethnic inequality in consumption expenditure in Indonesia. Hence, the present paper aims to contribute further to the discussion and debate on ethnicity and economic variables in the case of Indonesia, especially on the topic of polarization.

Recently, increasing attention has been paid to the measurement of income polarization (Duclos and Taptué 2015; Seshanna and Decornez 2003). The study of polarization covers social conflicts and social unrest (Chakravarty 2009; Esteban and Ray 1994), violent civil conflicts (Østby 2008), potential causes of rebellions and tensions (Pressman 2006), generators of tensions and social revolts (Alesina and Rodrik 1994), and ethnic conflicts (Forsberg 2008). The study of polarization is thus becoming essential in the context of social problems, mainly due to the existence of ethnic

conflicts. It is now generally acknowledged that polarization is more applicable than traditional inequality measurements to understanding levels of social tensions and conflict (Permanyer 2012). Inequality only measures the spread of welfare relative to the mean of the entire distribution. Polarization, by contrast, emphasizes the spread of welfare relative to the local means at several points along the distribution (Zhang and Kanbur 2001). Furthermore, the polarization measurement divides a population into significantly sized groups, such that members of a group feel very similar to other members of that group but very dissimilar to members of other groups (Esteban and Ray 1994). Thus, inequality and polarization do not always follow the same trends; they may move in the same direction simultaneously, but they can also move in opposite directions.

Our study makes a novel contribution by analyzing the effect of ethnicity on polarization. We investigate the effect of the ethnic group on earnings polarization in the Indonesian labour market using micro-level data. The dissertation applies Re-centered Influence Function (RIF) approach as initiated in Firpo et al. (2009) and in Fortin et al. (2011). Without incorporating any additional covariates, ethnicity has a significant effect on earnings polarization. However, when regional characteristics are included in the model, the effect of ethnicity on earnings polarization significantly decreases.

The rest of the paper is structured as follows. The next section presents the conceptual framework on ethnicity used in this study, contextual information on ethnic groups in Indonesia, and a basic outline of polarization as a concept. The third section explains the methodology, and the fourth section presents the data. The fifth section reports the results of the calculations. The final section concludes.

2 Conceptual framework

2.1 Defining ethnicity

The definition of ethnicity is still debated among cultural anthropologists, and the concept involves multiple explanations and theories. Baumann (2004) explored how previous scholars have defined ethnicity. The term *ethnicity* first appeared in the English Oxford Dictionary in 1953, and the term *ethnic* can be traced back to the term *ethnos*, a Greek term that relates to the words *band*, *tribe*, *race*, *people*, and *swarm* (Hutchinson and Smith 1996).

In previous studies, ethnicity has been conceptualized as a level of social stratification that incorporates race, class, kinship, age, estate, caste, and gender (Berreman 1981). He provided a clear distinction between race and ethnicity. In the first case, racial stratification relates to birth-ascribed status according to physical and cultural characteristics that people outside the group can observe. In the second case, ethnicity is associated with cultural characteristics defined within the ethnic group itself. Horowitz (1985) further widened this definition of ethnicity to describe the concept as a sense of collective belonging within a group due to common ancestry, language, historical moments, culture, race, caste, religion, or some combination of these factors.

More recently, Hutchinson and Smith (1996) proposed six main features of ethnicity: a common proper name, a common ancestry, shared historical memories, a common culture, a connection to the homeland, and a sense of solidarity. The common proper name is applied to distinguish an ethnicity from other communities. The myth of common ancestry builds the feeling of imagined relations due to a common origin in time and place. The shared historical memories express collective activities from the past, such as heroes, battles, and commemorations. The common culture generally encompasses belief/religion and language. The connection to homeland links

both people in the group who live in their homeland physically and those members who live outside their homeland (diaspora). Finally, the sense of solidarity builds strong relationships within the community.

Similarly, Jones (1997) defined three terms related to the concept of ethnicity: *ethnicity*, *ethnic identity*, and *ethnic group*. First, all the social and psychological phenomena related to a culturally constructed group identity can be described as ethnicity. Second, a self-identification within a specific group, as opposed to other groups, based on cultural or ancestry differences forms an ethnic identity. Third, any group that identifies apart from other groups based on cultural differences and/or shared descendants is considered as an ethnic group.

Finally, Baumann (2004) concluded that ethnicity is composed of self- and group identities as a combination of external factors (e.g., historical memories, sense of solidarity), internal factors (e.g., ancestry), and social interaction. Thus, ethnicity is an attributive symbol of an individual or group that can be reconstructed over time. In other words, ethnicity is not a static condition, but a fluid condition that can change over time depending on the circumstances. For instance, ethnic identity can change due to mixed marriages across ethnic groups, political changes, assimilation, and cultural exchanges (Baumann 2004; Ratcliffe 2010).

2.2 Ethnic groups and regional characteristics in Indonesia

Historically, before the independence of Indonesia, the Dutch East Indies colonial government segregated people based on their race. The European race was considered as the first class. The second class, known as the “foreign easterner group,” was immigrant groups from China or from Arab and other Asian regions. The third class was local ethnic groups (indigenous people) that had lived in the archipelago for at least a century, well before the arrival of immigrant groups. The census in 1932 conducted by the Dutch East Indies government recorded the population as 97 per cent indigenous people (*inlanders*), 2 per cent Chinese, 0.2 per cent other foreign easterners, and 0.4 per cent European people (Beets et al. 2002: 25).

Precise information on ethnic groups in Indonesia was not available from the 1940s until 1990s, mainly due to the treatment of public discussion of ethnicity as a political taboo (Ananta et al. 2014; Arifin et al. 2017). This regime had SARA policy, an acronym from ethnic (*suku*), religion (*agama*), race (*ras*), and inter-group (*antar-golongan*). The policy aimed to strengthen social and political stability and thereby to unify an ethnically diverse nation and overcome the ethnic discrimination experienced in the colonial era.

The central governments’ stance on ethnicity also prevented the Indonesian Statistical Office (BPS) from obtaining information about ethnic groups. Consequently, studies on ethnicity using micro-level data and quantitative approach in Indonesia are limited. To circumvent the lack of an ethnic group variable, some scholars have employed language spoken as the proxy of ethnicity, for instance Fearon (2003) and Alesina et al. (2003). After the disappearing of the SARA policy, the question about ethnicity was asked in the 2000 population census, then repeated in the 2010 population census.

Ananta et al. (2015) and Arifin et al. (2015) presented the statistics of ethnic groups in Indonesia according to the full data of the 2010 population census. Javanese and Sundanese ethnic groups are the first and the second largest ethnic groups, comprising 40 per cent and 16 per cent of the population, respectively. Both ethnic groups live mostly on Java island, which is the most densely populated island in Indonesia, yet they inhabit only 6 per cent of the land area. The other ethnic groups are Malay (3.7 per cent), Batak (3.6 per cent), Maduranese (3.0 per cent), Betawi (2.9 per cent), Minang (2.7 per cent), Buginese (2.7 per cent), and Bantenese (2.0 per cent). These groups

live mostly in specific regions in Indonesia. For instance, the Malay live in the Riau and Jambi provinces, the Batak in the North Sumatera province, the Maduranese on Madura island, and the Betawi in the Jakarta province. The rest of the population accounts for less than a quarter of the national population, which is divided into more than 600 ethnic groups.

Regarding the distribution of ethnic groups in each district, the Javanese comprise the largest ethnic group in 132 districts (out of 497 districts) (Arifin et al. 2015). Out of those 132 districts, 52 districts are outside of Java, with 39 districts in Sumatera and 8 districts in Kalimantan. However, none of these 52 districts are in Nusa Tenggara, Sulawesi, or Maluku regions. The highest percentages of the Javanese ethnic group are in Central Java, Yogyakarta, and East Java. The Malay ethnic group is the largest ethnic group in 34 districts in Indonesia. The Buginese and Dayak each comprise the largest ethnic groups in 22 districts. The Batak and Sundanese each comprise the largest ethnic group in 21 districts. Mostly, the provinces and districts in the Eastern Indonesia region have higher heterogeneity than those in Western Indonesia. The province of Central Java is home to the most homogeneous ethnic groups, predominantly the Javanese ethnic group, whereas the province of North Maluku is home to the most heterogeneous ethnic groups (Arifin et al. 2015).

In census and survey data, ethnic groups are usually defined based on the perception of the respondent (self-identification). However, a respondent's perception of his or her ethnic group may be fluid, and respondents can easily change their answer over time (Arifin et al. 2017). Chinese respondents are one example of the fluidity of ethnic identity over time. Chinese people initially came from various backgrounds, mostly Hokkien, Teochiu, Hakka, and Cantonese, in the southern region of modern China. The domination of the Chinese in the national economy increased the level of tension in the society in the 1990s. Thus, upon arriving in Indonesia during this period, Chinese people often hid their original ethnic identity, even changing their original Chinese names to more common Indonesian names. Since the democratisation era in 1998, though, Chinese ethnic groups have become more open to identifying themselves as Chinese (Arifin et al. 2017).

Another example of the fluidity of ethnic identity is the Dayak, an indigenous group on Kalimantan Island. Thung et al. (2004) recently reported an increase in open self-identification among the Dayak. Over the last century, the term *Dayak* was used to classify a tribal group, mostly referring to non-Muslim, non-Malay people living in forest areas covering parts of Indonesia and Malaysia. In recent years, identifying as part of the Dayak group has become a symbol of pride, uniting the indigenous, native people of the island. Currently, the classification of Dayak exists in the list of ethnic groups.

2.3 Earnings polarization

One of the most comprehensive literature on polarization measurements is the study of Duclos and Taptué (2015), which provided the origin of the concept of polarization and explained how polarization differs from inequality. Polarization, described as a simple phenomenon, is the shrinking and even disappearance of the middle class, which creates a more segregated society. One of the motivations for studying polarization is the ethical view that polarization enables understanding of distances and differences across groups, whereas inequality only compares differences across individuals. Much of the motivation to study polarization relates closely to the need to study social problems. In particular, those problems include conflicts and social unrest (Chakravarty 2009; Esteban and Ray 1994), violent civil conflicts (Østby 2008), potential causes of rebellions and tensions (Pressman 2006), generators of tensions and social revolts (Alesina and Rodrik 1994), and ethnic conflicts (Forsberg 2008).

The most influential concept of polarization was initiated by Esteban and Ray (1994), who introduced an identification-alienation framework. They argued that a high degree of homogeneity within each group (identification or internal homogeneity), a high degree of heterogeneity across groups (alienation or external heterogeneity), and a small number of significantly sized groups could increase income polarization.

Assume that the income variable (or another welfare indicator, such as earnings, expenditures, and assets) can be split into a finite number of groups $i = 1, 2, \dots, n$, where each group has an income γ_i . An identification function can be formulated as an increasing function $I(n_i)$, where n_i is the number of individuals in group i . The distance between individual i and individual j is formulated as $\delta(\gamma_i, \gamma_j)$, where γ_i and γ_j are the incomes of individuals i and j , consecutively. Individual i feels alienation $a[\delta(\gamma_i, \gamma_j)]$ towards individual j .

However, the approach above is only applicable for a discrete variable, where the number and locations of the income groups are assumed to be arbitrary. Thus, Duclos et al. (2004) extended this approach for a continuous variable:

$$DER = \iint T[f(x), |x - y|]f(x)f(y) dx dy. \quad (1)$$

The equation above is known as the Duclos-Esteban-Ray (DER) index. It reflects the sum of all effective antagonism over a continuous variable. The identification function is derived from $f(x)$ as a non-normalized density function, whereas $f(y)$ is another density function. The alienation function is derived from $a = |x - y|$ as the distance between an individual with income x and an individual with income y . Therefore, $T[f(x), |x - y|]$ is an increasing function of effective antagonism.

The measurement of polarization is proportional to the equation

$$DER(f, \alpha) \equiv \iint f(x)^{1+\alpha}f(y)|x - y| dx dy. \quad (2)$$

The DER index is widely used in polarization cases. The alienation component can be formulated as $\bar{a} = \iint |x - y| dF(x) dF(y)$, and the identification component can be written as $\bar{i} = \int f(x)^{1+\alpha} dx$. A higher value of α implies a higher level of within-group identification in the formula and larger differences in the inequality index. A higher value of α therefore expresses a stronger homogeneity among the individuals within a group. The DER index becomes the Gini index when $\alpha = 0$. To fulfill the axioms in their formula, α should be bounded, $\alpha \in [0.25, 1]$. Our study employs the value of $\alpha = 0.5$ as the common application. We also imply a set different value of α for a sensitivity procedure.

The DER index is constructed using the Distributive Analysis Stata Package (DASP) that was developed by Abdelkrim and Duclos (2007). DASP has been used widely to perform distribution analyses, such as analyses of poverty, inequality, decomposition by subgroups of the population, decomposition by income sources, poverty elasticity, polarization, and benefit incidence. Specifically, we use the command `ipolder` from DASP v2.3. The result of the calculation includes an alienation component, an identification component, the correlation between the alienation and identification components, a total estimation of the DER index, and standard error.

3 Methodology

This section presents the detailed methodology used to accomplish the objectives of this study. As explained in the introduction, we aim to investigate the effect of the ethnic group on earnings polarization. F is the distribution function of earnings, and $v(F)$ is a functional statistic of distribution, for instance, the mean, variance, quantile, inequality, and polarization. We use the DER index as a functional statistic of distribution in this study. Then, F can be denoted as the combination of the sub-group earnings distribution for different ethnicities (ethnic groups):

$$F(y) = \sum_{x \in \mathbb{C}_x} \pi_x F_x(y), \quad (3)$$

where F_x is the earnings distribution among workers of an ethnic group of x , π_x is the proportion of workers of an ethnic group of x in the population, and \mathbb{C}_x is the set of K workers type.

We attempt to calculate the effect of a marginal substitution of the ethnic reference group by other ethnic groups on $v(F)$. We choose Javanese, the largest ethnic group in Indonesia, as the reference group. This method is known as an unconditional partial effect/UPE (Firpo et al. 2009), a policy effect (Rothe 2010), and a counterfactual effect (Chernozhukov et al. 2013). Following Choe and Van Kerm (2014), the formal equation can be written as follows:

$$UPE(v(F), k) = \lim_{t \rightarrow 0} \frac{v(G_r^{F,t,k}) - v(F)}{t} \quad (4)$$

where $G_r^{F,t,k}$ is the earnings distribution after substituting a proportion t of reference workers of ethnic r (Javanese) for workers from ethnic k , as written below:

$$G_r^{F,t,k}(y) = (\pi_k + t)F_k(y) + (\pi_r - t)F_r(y) + \sum_{x \in \mathbb{C}_x \setminus \{k,r\}} \pi_x F_x(y). \quad (5)$$

The equation above explains the earnings distribution after the substitution of a part of the reference group by other ethnic groups.

The equation of $UPE(v(F), k)$ equals to

$$UPE(v(F), k) = \int RIF(y; v, F) d(G_r^{F,t,k} - F)(y). \quad (6)$$

$RIF(y; v, F)$ is the Re-centered Influence Function (RIF). The RIF is constructed by adding the influence function to the function itself. According to Hampel (1974), the influence function is defined as

$$IF(y; v, F) = \lim_{\epsilon \rightarrow 0} \frac{v((1-\epsilon)F + \epsilon \Delta y) - v(F)}{\epsilon}. \quad (7)$$

The influence function explains the effect on $v(F)$ of an infinitesimal contamination of F at the point mass y (Essama-Nssah and Lambert 2012).

The effect of an ethnic group on a functional statistic of distribution can be expressed as follows:

$$UPE(v(F), k) = E[RIF(y; v, F)|X = k] - E[RIF(y; v, F)|X = r], \quad (8)$$

where $E[RIF(y; v, F)|X = k]$ is the expected value of $RIF(y; v, F)$ from ethnic k , and $E[RIF(y; v, F)|X = r]$ is the expected value of $RIF(y; v, F)$ from Javanese ethnic group.

However, the equation above neglects individual characteristics beyond ethnic groups. Individual characteristics may include, for example, level of education, working experience, and age. To address this issue, we consider a marginal substitution between the reference group and other ethnic groups that are conditional on their characteristics. The effect of ethnic group substitution is referred to as a conditional unconditional partial effect (CUPE):

$$CUPE(v(F), k) = \lim_{t \rightarrow 0} \frac{v(H_r^{F,t,k}) - v(F)}{t}, \quad (9)$$

where $H_r^{F,t,k}$ is the earnings distribution after substituting a proportion t of reference workers of ethnic r (Javanese) for workers of ethnic k .

$$H_r^{F,t,k}(y) = \int_{\mathbb{C}_Z} \left((\pi_{k|z} + t)F_{k,z}(y) + (\pi_{r|z} - t)F_{r,z}(y) + \sum_{x \in \mathbb{C}_x \setminus \{k,r\}} \pi_{x|z} F_{x,z}(y) \right) f_Z(z) dz \quad (10)$$

$F_{x,z}(y)$ represents the conditional earnings distribution given worker ethnic $X = x$ and characteristics $Z = z$. Furthermore, $\pi_{k|z}$ denotes the share of workers of ethnic x among workers with characteristics $Z = z$.

Per the above expression on UPE, the effect of an ethnic group on a functional statistic of distribution can be expressed as follows:

$$CUPE(v(F), k) = \int_{\mathbb{C}_Z} E[RIF(y; v, F)|X = k, Z = z] - E[RIF(y; v, F)|X = r, Z = z] f_Z(z) dz. \quad (11)$$

To answer the research question, we apply three-step estimations of RIF-regression. First, we compute the value of RIF for each observation i for the statistic of interest $v(F)$. Following equation (7), the influence function of observation i can be calculated by measuring the gap between the modified statistic of interest (with the individual i) and the initial statistics of interest (without the individual i), then divided by $1/N$ ($\varepsilon \rightarrow 0$). The value of RIF equals to the sum of the initial statistics of interest and the influence function. Second, $RIF(y_i; v, F)$ is regressed by the common OLS and excluding other covariates as follows:

$$E[RIF(y; v, F)|X = x, Z = z] = \alpha + x\beta. \quad (12)$$

x is a vector of ethnic groups, whereas α is the constant representing any variables that cannot be captured in the model. The effect of an ethnic group can be obtained from the estimated coefficients of β , where Javanese is defined as the reference group (omitted variable).

Third, we regress $RIF(y_i; v, F)$ including other covariates as follows:

$$E[RIF(y; v, F)|X = x, Z = z] = \alpha + x\beta + z\gamma. \quad (13)$$

Again, the effect of an ethnic group can be obtained from the estimated coefficients of β by including individual workers characteristics z .

4 Data

The primary data source in this study is the Indonesian Family Life Survey (IFLS), which offers rich micro-level data that combine ethnicity and socio-economic variables. In particular, we use the IFLS dataset in 2014/2015 (hereafter written as 2015), conducted by the RAND Corporation (California, the US) in collaboration with the Survey METER (Yogyakarta, Indonesia). The ethnic group becomes the main proxy of ethnicity, whereas individual monthly income represents an economic variable. An ethnic group is a self-reported and open question in the survey, so the respondent can answer this question without having to select from a list of choices of ethnic groups.

The following filters were applied in the dataset. The samples are restricted to all workers aged 15–60 who have positive earnings and are not engaged in any educational institution. We include only male-headed households to avoid the potential effect of household structure. We exclude unpaid family workers who have implicit earnings or high unreported earnings and who simply reported difficulty in evaluating their earnings. The government workers are excluded from our observations due to the specific institutional and earnings regulations that apply to the government workers. Most importantly, we only analyze the samples with reported ethnic groups. Thus, the selection leaves a sample size of 14,370 observations from self-employed, private sector workers, casual workers in agriculture, and casual workers in non-agriculture.

In this study, we classify potential individual control variables into the labour market, human capital, demographic, sectoral, and spatial characteristics. Labour market characteristics include working hours and occupation category. Human capital characteristics are represented in working experiences and level of education. Demographic components cover gender, marital status, and age of the individual. The sectoral variable is the classification of firm or company into primary (agriculture and mining), secondary (manufacture, electricity, gas, water, and construction), and tertiary sectors (wholesale, retail, restaurant, hotel, transportation, storage, communication, finance, insurance, real estate, and social services). Spatial characteristics are determined by where the individual lives, namely whether urban or rural and in which region (province).

Table 1 Characteristics across ethnic groups

Ethnic	Fraction (%)	Earnings (IDR 1,000)	Female (%)	Years of education	Farmer (%)	Type of employment		Urban (%)
						Self-employed (%)	Casual workers (%)	
Javanese	54.5	1,454	34.3	9.3	22.7	40.4	14.5	50.7
Sundanese	15.4	1,568	30.5	9.2	14.6	40.5	13.0	65.4
Maduranese	4.0	949	40.6	8.0	17.6	41.2	21.3	38.6
Palembang, Melayu, & oth Sumaterans	3.9	1,909	26.1	9.8	27.9	45.7	11.8	52.2
Betawi	3.8	2,172	32.0	9.7	8.2	32.2	9.8	89.9
Bugis, Makassar, & Toraja	2.9	1,417	30.4	8.7	30.9	56.4	11.0	46.7
Batak & Nias	2.8	1,403	28.9	10.4	38.8	58.8	11.0	37.3
Minang	2.6	1,663	31.8	10.4	22.9	50.1	15.6	52.7
Sasak, Bima-Dompu, & Sumbawa	2.4	1,229	28.7	8.8	31.9	51.7	14.8	49.6
Bali	2.0	1,647	43.0	9.3	18.2	47.2	10.2	58.6
Chinese	0.43	2,729	27.3	12.5	2.4	41.0	4.8	97.2
Others	5.3	1,397	32.2	9.0	17.8	46.5	10.3	65.6
Average		1,501	33.1	9.3	21.5	42.3	13.8	54.7

Note: Type of employment consists of self-employed, private sector workers (omitted), and casual workers.

Source: The authors' calculation based on IFLS 5.

Table 1 displays individual characteristics across a number of ethnic groups in Indonesia. Based on IFLS classification, samples are classified into 12 ethnic groups. Classification has been determined based on the location of each ethnic group. For instance, members of the Sundanese ethnic group live mostly in the West Java province. Members of the Sasak, Bima-Dompu, and Sumbawa groups live mostly in the West Nusa Tenggara province. Over half of the samples (55 per cent) come from the Javanese ethnic group, followed by the Sundanese ethnic group (15 per cent). The remaining ethnic groups are mainly the Maduranese (4 per cent), Palembang, Melayu, and other Sumaterans (4 per cent), and Betawi (4 per cent) groups. Other ethnicities comprise less than 3 per cent of the sample. Javanese workers are dominant in the sample due to the nature of IFLS. Based on its survey design, IFLS was first conducted in 1993 in 13 of Indonesia's 27 provinces, mostly in the western region. Although the survey covered less than half of the provinces, IFLS represented around 83 per cent of the Indonesian population.

In terms of education, there are significant differences in earnings across ethnic groups in Indonesia. Chinese workers have the highest earnings, although their share of the population is small. Individual earnings are lowest among Maduranese workers and highest among Chinese workers. Chinese workers earn nearly 90 per cent more than Javanese workers, who constitute the majority of the Indonesian population. Compared to the Sundanese ethnic group, Chinese workers earn almost 80 per cent higher wages. In terms of the lowest earnings, Maduranese workers earn only one-third of the wages of Chinese workers.

Table 2 Dominant Ethnic Fraction

Panel A: Dominant ethnic fraction		
Province	%	Contribution
North Sumatera	4.6%	Batak & Nias 52.2%, Javanese 37.7%.
West Sumatera	2.8%	Minang 82.2%.
South Sumatera & Bangka Belitung	3.8%	Palembang, Melayu, & oth Sumaterans 56.9%, Javanese 33.9%.
Lampung	3.7%	Javanese 61.2%, Sundanese 17.5%, Palembang, Melayu, & oth Sumaterans 14.8%.
Jakarta	6.0%	Betawi 34.7%, Javanese 30.5%, Sundanese 21.5%.
West Java	18.7%	Sundanese 61.8%, Javanese 20.4%.
Central Java	17.2%	Javanese 97.6%.
Yogyakarta	6.2%	Javanese 98.8%.
East Java	21.8%	Javanese 81.4%, Maduranese 17.2%.
Banten	5.7%	Javanese 39.5%, Sundanese 25.4%, Betawi 16.7%, Palembang, Melayu, & oth Sumaterans 10.6%
Bali	2.1%	Bali 86.8%.
West Nusa Tenggara	2.3%	Sasak, Bima-Dompu, & Sumbawa 96.7%
South Kalimantan	2.6%	Banjar 70.6%, Javanese 17.9%,
South & West Sulawesi	2.6%	Bugis, Makassar, & Toraja 95.2%.
Panel B: The location of ethnic		
Ethnic group	%	Contribution
Javanese	54.5%	East Java 32.6%, Central Java 30.7%, Yogyakarta 11.3%, West Java 7%, Banten 4.1%, Lampung 4.1%, Jakarta 3.4%.
Sundanese	15.4%	West Java 75.2%, Banten 9.4%, Jakarta 8.4%.
Maduranese	4.0%	East Java 93.6%.
Palembang, Melayu, & oth Sumaterans	3.9%	South Sumatera & Bangka Belitung 54.7%, Banten 15.4%, Lampung 14.0%.
Betawi	3.8%	Jakarta 54.6%, Banten 24.8%, West Java 19.4%.
Bugis, Makassar, & Toraja	2.9%	South & West Sulawesi 82.9%.
Batak & Nias	2.8%	North Sumatera 86.4%.
Minang	2.6%	West Sumatera 87.9%.
Sasak, Bima-Dompu, & Sumbawa	2.4%	West Nusa Tenggara 94.1%
Bali	2.0%	Bali 92.5%.
Chinese	0.43%	Jakarta 23.4%, North Sumatera 20.9%, East Java 15.4%, West Sumatera 12.2%, West Java 9.3%, South Sumatera & Bangka Belitung 8.4%.

Source: The authors' calculation based on IFLS 5.

Female participation is relatively high, ranging from 25 per cent (Palembang, Melayu, and other Sumaterans) to over 40 per cent (Maduranese and Bali). Significant differences occur in education characteristics across ethnic groups, where Chinese workers over 12 years of education and have completed high school education. This ethnic group is then followed by Minang, Bugis, Makassar, Toraja, Palembang, Melayu, other Sumaterans, and Betawi, who have 10 years of education. The remaining ethnic groups have around 9 years of education, equivalent to a junior high school level.

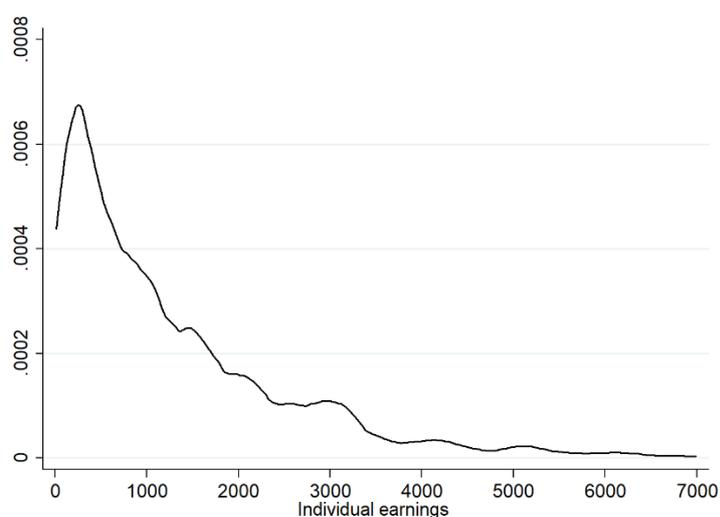
Occupation classification differs significantly across ethnic groups. A small percentage of Chinese workers are farmers (2.4 per cent), whereas the percentage of farmers across the remaining ethnic groups ranges from 8 per cent (Betawi) to 32 per cent (Sasak, Bima-Dompu, and Sumbawa). The low percentage of farmers among Chinese and Betawi workers may be due to the high percentage of workers living in urban areas, 97 per cent and 90 per cent, respectively. On average, 42 per cent of workers are self-employed, 14 per cent of them are casual workers, and the remaining 46 per cent of workers are employed in the private sector.

The descriptive data shows that each ethnic group lives in particular regions (Table 2). For instance, the Javanese ethnic group lives mostly in the Java region, such as East Java, Central Java, and Yogyakarta. Three quarters of the Javanese ethnic group live in these regions, which are dominated by the Javanese ethnic group. 98 and 99 per cent of people living in Central Java and Yogyakarta are from the Javanese ethnic group, respectively. In East Java, the Javanese and Maduranese ethnic groups constitute 81 and 17 per cent of the population, respectively. Three quarters of the Sundanese ethnic group live in West Java. West Java is the origin of the Sundanese ethnic group and hosts sixty per cent of the population. The second largest ethnic group in West Java is Javanese, which constitutes two tenths of the population.

5 Results

This section discusses the results of our calculation. First, the section discusses the effect of ethnic groups on earnings polarization, including the effect of regional characteristics, using the RIF-regression method. A sensitivity analysis is then conducted by analyzing the interaction effect, calculating various polarization indices, changing the reference group, and re-classifying ethnic groups.

Figure 1 Kernel Density of Individual Earnings



Note: All values are in thousand Indonesian Rupiah (1,000 IDR) in July 2015. The calculation is weighted.

Source: The authors' illustration based on IFLS 5.

Figure 1 displays the kernel density of individual earnings in Indonesia. Based on a visual inspection of the graph, there is a significant pole in the distribution at approximately IDR300,000-350,000. A middle significant pole appears at approximately IDR2,800,000-3,200,000. A number of smaller poles are detected at approximately IDR1,500,000, IDR2,000,000, IDR4,200,000, and IDR5,200,000. Individuals at those multiple earnings peaks may feel similarity (identification) with one another and a sense of dissimilarity from other earnings groups (alienation). The clustering of earnings at some points of distribution may improve earnings polarization.

Table 3 Estimated coefficients of RIF-regression

Coefficient	UPE		CUPE		
	Model 1	Model 2	Model 3	Model 4	Model 5
Ethnic group					
Javanese	Omitted	Omitted	Omitted	Omitted	Omitted
Sundanese	0.058**	0.044**	0.012	-0.046**	-0.133+
Maduranese	-0.248**	-0.253**	-0.230**	-0.056**	-0.076
Palembang, Melayu, oth Sumaterans	0.275**	0.247**	0.241**	0.020	-0.267
Betawi	0.145**	0.123**	0.038+	-0.027	-0.025
Bugis, Makassar, Toraja	0.285**	0.281**	0.277**	-0.006	0.180
Batak, Nias	0.266**	0.236**	0.261**	0.072*	-0.108
Minang	0.276**	0.245**	0.244**	0.047	-0.064
Sasak, Bima-Dompu, Sumbawa	0.286**	0.291**	0.279**	0.081	0.281
Bali	0.260**	0.251**	0.237**	0.054	0.179
Chinese	0.140*	0.103	0.039	-0.070	-0.013
Other ethnic groups	0.216**	0.212**	0.177**	-0.026	-0.041
Labour market characteristics	NA	Yes	Yes	Yes	Yes
Human capital characteristics	NA	Yes	Yes	Yes	Yes
Demographic characteristics	NA	Yes	Yes	Yes	Yes
Sector	NA	Yes	Yes	Yes	Yes
Urban	NA	NA	Yes	Yes	Yes
Province	NA	NA	NA		
North Sumatera				0.170**	0.166**
West Sumatera				0.157**	0.144
South Sumatera & Bangka Belitung				0.241**	0.288**
Lampung				0.308**	0.343**
Jakarta				0.027	-0.013
West Java				-0.010	-0.014
Central Java				Omitted	Omitted
Yogyakarta				0.045*	-0.042*

East Java				-0.245**	-0.251**
Banten				-0.033	-0.050**
Bali				0.132*	0.016
West Nusa Tenggara				0.161 ⁺	-0.092
South Kalimantan				0.222**	0.224**
South & West Sulawesi				0.250**	0.083
Interaction: Ethnic group & province	NA	NA	NA	NA	Yes
Ethnicity					
F-value	79.87	73.79	71.23	2.86	0.56
p-value	<0.001	<0.001	<0.001	<0.001	0.86
Province					
F-value	NA	NA	NA	81.42	72.93
p-value	NA	NA	NA	<0.001	<0.001
Interaction: Ethnic group & province					
F-value	NA	NA	NA	NA	15.94
p-value	NA	NA	NA	NA	<0.001

Note: NA: Not applicable. We use $\alpha = 0.50$ for calculating earnings polarization. Model 1 uses no additional control variables (UPE). Model 2 – model 4 include additional control variables (CUPE). Model 2 captures labor market characteristics (working hours per week, occupation classification, and employment category), human capital (working experience and education level), demographic characteristics (gender, marital status, and age), and sector (primary, secondary, or services). Model 3 covers urban/rural residential location as additional explanatory variables to Model 2. Model 4 covers regional characteristics (province) as additional explanatory variables to Model 3. Model 5 captures the interaction effect of ethnic group and province as additional explanatory variables.

The primary sector includes agriculture and mining. The secondary sector includes manufacture, electricity, gas, water, and construction. Services sector includes wholesale, retail, restaurant, hotel, transportation, storage, communication, finance, insurance, real estate, and social services.

Source: The authors' calculation based on IFLS 5.

The results in Table 3 presented the estimated RIF-regression coefficient. Using Javanese ethnicity as a reference, and without additional control variables (covariates) in Model 1 (UPE), the estimated coefficients are largely positive and significant. This finding suggests that non-Javanese workers tend to increase earnings polarization significantly (0.06 to 0.29), except for the Maduranese ethnic group (-0.25). A higher polarization index is associated with a higher probability of social problems. Similarly, a lower polarization index is associated with a lower probability social conflict.

However, it is also useful to consider additional control variables (covariates), which may also be referred to as CUPE. Model 2 incorporates the labor market, human capital, demographics, and sectoral characteristics as additional control variables. The Javanese ethnic group is still the reference group (omitted variable) in this equation. The majority of ethnic groups tend to increase earnings polarization significantly, but the effects of the ethnic group are less significant than those in Model 1. For instance, in Model 1 and Model 2, the effect of the Betawi ethnic group is 0.15 and 0.12, respectively. However, while the Chinese ethnic group no longer has a significant effect on earnings polarization, the Maduranese ethnicity still has an adverse effect on earnings polarization (polarization-decreasing).

When incorporating urban/rural residential location (Model 3), the effect of non-Javanese ethnic groups is less significant. For instance, Sundanese, Palembang, Melayu, other Sumaterans, Bugis, Makassar, Toraja, Minang, Sasak, Bima-Dompu, Sumbawa, and Bali have a less significant effect than in Model 2. Betawi and Chinese ethnic groups no longer have a significant effect on earnings polarization.

Ultimately, the results in Model 4, which uses regional characteristics, differ from those in model 3. Regional characteristics have a significant effect on polarization, reducing the effect of ethnicity on earnings polarization. The value of F-statistics of ethnicity decreases significantly from Model 3 to Model 4. All observations are classified into 14 provinces, and Central Java is an omitted

variable. All provinces have a significant effect on earnings polarization, except for Jakarta, West Java, Banten, and West Nusa Tenggara. Just two of the 14 ethnic groups have a significant effect on earnings polarization, namely the Maduranese ethnic group (polarization-decreasing) and the Batak-Nias ethnic group (polarization-increasing).

Ethnic group may be correlated with province because each ethnic group tends to live in a specific province. For instance, most Batak and Nias groups live in North Sumatera, most Balinese groups live on Bali island, and most Javanese groups live on Java island. To incorporate the correlation between ethnicity and province, a new equation is estimated using an interaction variable between ethnic group and regional characteristics in a sensitivity analysis (Model 5). According to our calculation, the interaction effect is significant in the equation (F-statistics is 15.94 and p-value is less than 0.001).

However, ethnic groups and regional characteristics have a different pattern. Ethnic groups become not significant to the equation (p-value = 0.86). The value of F-statistics of ethnicity declines significantly from 2.86 to 0.56. Regional characteristics have a significant effect earnings polarization (F-statistics is 72.93 and p-value is less than 0.001). A number of provinces have a significant effect on earnings polarization, namely North Sumatera, South Sumatera, Bangka Belitung, Lampung, Yogyakarta, East Java, Banten, and South Kalimantan. This finding implies that, after incorporating the interaction effect, ethnicity becomes not relevant to the equation. Moreover, the effect of regional characteristics is stronger than the effect of ethnic groups, indicated by the larger value of F-statistics.

In the first sensitivity analysis (Appendix: Table 8A.1), the reference ethnic group is changed from Javanese to Maduranese (the group with the lowest average earnings) and Sundanese (the second largest of ethnic groups). This analysis produces similar findings. Without incorporating the interaction effect (between ethnic group and province), ethnic groups have a significant effect on earnings polarization and regional characteristics have a stronger effect than ethnic group. By incorporating the interaction effect, ethnicity has no longer significant effect and regional characteristics has significant effect to earnings polarization.

In the second sensitivity analysis, the results are robust to varying degrees of identification in the polarization indices (Appendix: Table 8A.2). The effect of ethnic groups may be positive or negative, depending on the degree of identification component (α). The findings suggest that the effect of ethnic groups is less significant on earnings polarization when incorporating regional characteristics into the model. We have a similar conclusion that ethnicity becomes insignificant when the model captures the interaction effect. Additionally, regional characteristics has a significant effect to earnings polarization.

In the final robustness check, the calculation is replicated with a different classification of ethnic groups. In the beginning, the samples are classified into 12 ethnic groups. These samples are then re-classified into 7 categories, namely (1) Javanese, (2) Sundanese, Betawi, (3) Maduranese, (4) Batak, Nias, Minang, Palembang, Melayu, other Sumaterans, (5) Bali, Sasak, Bima-Dompu, Sumbawa, (6) Bugis, Makassar, Toraja, and (7) other ethnic groups. Sundanese and Betawi ethnic groups are located mostly in Jakarta, West Java, and Banten. The Maduranese ethnic group is mostly located on Madura island, a region of East Java. Javanese and Maduranese groups are not combined into one ethnic group in order to distinguish the effect of each group. Batak, Nias, Minang, Palembang, and Melayu are located mostly in the Sumatera region. Bali, Sasak, Bima-Dompu, and Sumbawa are located mostly in the Bali-Nusa Tenggara region. Using the Javanese group as reference, the results are similar to those of the original model (Appendix: Table A.3, Panel A). By incorporating additional explanatory variables, non-Javanese ethnic group still have a polarization-enhancing effect, but this effect is smaller than in the previous estimation. However, non-Javanese ethnic groups do not have a significant on earnings polarization when regional

characteristics are considered. Therefore, the effect of ethnic groups is less significant when incorporating regional characteristics.

Then, the samples are re-classified into three ethnic groups for the robustness check, namely (1) Javanese, (2) Sundanese and Betawi, and (3) other ethnic groups (Appendix: Table 8A.3, Panel B). Ultimately, the samples are classified only into two ethnic groups (Appendix: Table 8A.3, Panel C). We found the similar conclusion in three panels, where the effect of ethnicity becomes not significant, but the effect of regional characteristics is still significant by incorporating the interaction effect. The interaction effect covers a cross-interaction between ethnicity and regional.

6 Concluding remarks

In this study, a RIF-regression is used to analyze the effect of ethnicity on earnings polarization. While the applicability of the results is limited to Indonesia, our research contributes to the literature on ethnicity and earnings distribution (labour market). Existing literature on ethnicity in Indonesia has largely focused on ethnicity itself without considering economic variables in the analyses. This research enriches the discussion on ethnicity in Indonesia by analyzing earnings polarization and explanatory variables, such as the effect of regional characteristics.

This paper used the latest available dataset from the Indonesian Family Life Survey in 2014/2015, which covered ethnic group and earnings variables. The RIF value of the DER polarization index was regressed using the ethnic group as the explanatory variable. Following this regression, additional explanatory variables/covariates, including regional characteristics, were considered. Without additional covariates, ethnic group has a significant effect on earnings polarization. However, the effect of ethnic group becomes less significant when including additional covariates, such as labour market structure, human capital condition, demographic characteristics, and sectoral classification. Finally, the effect of ethnic group becomes even less significant when incorporating regional characteristics into the estimation.

According to the findings of this paper, ethnicity (ethnic diversity) is becoming less and less significant in earnings polarization when incorporating regional characteristics. We found that regional characteristics are more significant than ethnic characteristics. The lessening impact of ethnic group is caused by the strong correlation between ethnic group and regional characteristics. For instance, the majority of the Javanese group live on Java island, the majority of the Sundanese group live in the province of West Java, the majority of the Batak group live in the province of North Sumatera, and the majority of the Balinese group live in the province of Bali (Table 8.2). Thus, by incorporating the interaction effect, ethnic groups become irrelevant on earnings polarization. The robustness of this paper's findings has been ensured by changing the reference ethnic group, estimating the varying degrees of DER polarization index, and re-classifying the ethnic groups.

Several policy recommendations can be provided based on the findings of this paper. Firstly, policy makers may focus on reducing the development gap across regions and islands. Indonesia is a large archipelagic country with 17,000 islands. Connectivity and infrastructure development should be the primary goals of development. Secondly, the government of Indonesia has recently introduced a village budget (*dana desa*) to encourage rural development. This budget may reduce the development gap between urban and rural areas. However, the budget should be allocated to the proper programs, such as those that aim to improve village infrastructure, build village-level companies to reduce unemployment rates, and implement vocational training schemes.

References

- Abdelkrim, A. and J-Y. Duclos (2007). 'DASP: Distributive Analysis Stata Package', (PEP, The World Bank, UNDP and Université Laval).
- Alesina, A. and D. Rodriki (1994). 'Distributive Politics and Economic Growth', *Quarterly Journal of Economics*, 109, 465-89.
- Alesina, A., R. Baqir, and W. Easterly (1999). 'Public goods and ethnic divisions', *The Quarterly Journal of Economics*, 114 (4), 1243-84.
- Alesina, A., C. Gennaioli, and S. Lovo (2014). 'Public goods and ethnic diversity: Evidence from deforestation in Indonesia'. No 20504, *NBER Working Papers*. Cambridge, MA: National Bureau of Economic Research, Inc.
- Alesina, A. et al. (2003). 'Fractionalization', *Journal of Economic growth*, 8 (2), 155-94.
- Ananta, A. et al. (2014). 'A New Classification of Indonesia's Ethnic Groups (Based on the 2010 Population Census)', *ISEAS Working Papers #1*. Singapore: Institute of Southeast Asian Studies
- Ananta, A., et al. (2015). *Demography of Indonesia's ethnicity*. Singapore: Institute of Southeast Asian Studies.
- Arifin, E.N. , M.S. Hasbullah, and A. Pramono (2017). 'Chinese Indonesians: How many, who and where?', *Asian Ethnicity*, 18 (3), 310-29.
- Arifin, E.N. et al. (2015). 'Quantifying Indonesia's ethnic diversity: Statistics at national, provincial, and district levels', *Asian Population Studies*, 11 (3), 233-56.
- Baumann, T. (2004). 'Defining ethnicity', *The SAA archaeological record*, 4 (4), 12-14.
- Beets, G. et al. (2002). 'De demografische geschiedenis van de Indische Nederlanders', *Den Haag: NIDI*, 59-65.
- Berreman, G. (1981). 'Social inequality: Across-cultural approach', *Social inequality: Comparative and developmental approaches*. New York: Academic Press.
- Bertrand, J. (2004). *Nationalism and ethnic conflict in Indonesia*. Cambridge University Press.
- Chakravarty, S.R. (2009). *Inequality, polarization, and poverty: Advances in distributional analysis* [online text], Springer.
- Chernozhukov, V., I. Fernández-Val, and B. Melly (2013). 'Inference on counterfactual distributions', *Econometrica*, 81 (6), 2205-68.
- Choe, C. and Ph. van Kerm (2014). 'Foreign workers and the wage distribution: Where do they fit in?', *LISER working papers*. Luxembourg: Luxembourg Institute of Socio-Economic Research.
- Duclos, J-Y. and A-M. Taptué (2015). 'Polarization', *Handbook of income distribution* (2: Elsevier), 301-58.
- Duclos, J-Y., J. Esteban, and D. Ray (2004). 'Polarization: Concepts, measurement, estimation', *Econometrica*, 72 (6), 1737-72.
- Essama-Nssah, B. and P.J. Lambert (2012). 'Influence Functions for Policy Impact Analysis', *Inequality, Mobility and Segregation: Essays in Honor of Jacques Silber*, 135-59.
- Esteban, J. and D. Ray (1994). 'On the measurement of polarization', *Econometrica*, 62 (4), 819-51.

- Esteban, J. and D. Ray (2008). 'Polarization, fractionalization and conflict', *Journal of Peace Research*, 45 (2), 163-82.
- Esteban, J. and G. Schneider (2008). 'Polarization and conflict: Theoretical and empirical issues', *Journal of Peace Research*, 45 (2), 131-41.
- Esteban, J. L. Mayoral, and D. Ray (2012). 'Ethnicity and conflict: An empirical study', *American Economic Review*, 102 (4), 1310-42.
- Fearon, J.D. (2003). 'Ethnic and cultural diversity by country', *Journal of Economic Growth*, 8 (2), 195-222.
- Firpo, S.N.M. Fortin, and T. Lemieux (2009). 'Unconditional quantile regressions', *Econometrica*, 77 (3), 953-73.
- Forsberg, E. (2008). 'Polarization and ethnic conflict in a widened strategic setting', *Journal of Peace Research*, 45 (2), 283-300.
- Fortin, N.M., T. Lemieux, and S. Firpo (2011). 'Decomposition methods in economics', in Orley Ashenfelter and David Card (eds.), *Handbook of labor economics* (4A; Amsterdam: Elsevier).
- Habyarimana, J. et al. (2007). 'Why does ethnic diversity undermine public goods provision?', *American Political Science Review*, 101 (4), 709-25.
- Hampel, F.R. (1974). 'The influence curve and its role in robust estimation', *Journal of the American Statistical Association*, 69 (346), 383-93.
- Horowitz, D.L. (1985). *Ethnic groups in conflict* (University of California Press).
- Hutchinson, J. and A.D. Smith (1996). 'Introduction', in John Hutchinson and Anthony D Smith (eds.), *Ethnicity* (Oxford and New York: Oxford University Press), 3-16.
- Jones, S. (1997). *The archaeology of ethnicity: Constructing identities in the past and present* (Psychology Press).
- Merkel, W. and B. Weiffen (2012). 'Does heterogeneity hinder democracy?', *Comparative Sociology*, 11 (3), 387-421.
- Muller, C. (2016). 'Ethnic inequality and community activities in Indonesia'. *WIDER Working Paper 2016/70*. Helsinki: UNU-WIDER.
- Østby, G. (2008). 'Polarization, horizontal inequalities and violent civil conflict', *Journal of Peace Research*, 45 (2), 143-62.
- Permanyer, I. (2012). 'The conceptualization and measurement of social polarization', *The Journal of Economic Inequality*, 10 (1), 45-74.
- Pressman, S. (2006). 'The Decline of the Middle Class: An International Perspective', *Luxembourg Income Study Working Paper Series No. 280*.
- Ratcliffe, P. (2010). 'Ethnic group', *Sociopedia. isa*, 1-10.
- Rothe, C. (2010). 'Nonparametric estimation of distributional policy effects', *Journal of Econometrics*, 155 (1), 56-70.
- Schündeln, M. (2013). 'Ethnic heterogeneity and the private provision of public goods', *The Journal of Development Studies*, 49 (1), 36-55.
- Seshanna, S. and S. Decornez (2003). 'Income polarization and inequality across countries: An empirical study', *Journal of Policy Modeling*, 25 (4), 335-58.

- Thung, J.-L., Y. Maunati, and P.M. Kedit (2004). *The (re)construction of the 'Pan Dayak' identity in Kalimantan and Sarawak: A study on minority's identity, ethnicity, and nationality*. Jakarta: Puslit Kemasyarakatan dan Kebudayaan LIPI.
- Zhang, X. and R. Kanbur (2001). 'What difference do polarisation measures make? An application to China', *Journal of Development studies*, 37 (3), 85-98.

Appendix

Table A.1 presents the result of sensitivity analysis by changing the reference of an ethnic group. The first three columns exercise Maduranese ethnic group as the reference, whereas the next three columns use Sundanese ethnic group as the reference. Model 3 covers labour market (working hours per week, occupation classification, and employment category), human capital (working experience and education level), demographic characteristics (gender, marital status, and age), sector (primary, secondary, or services), and urban/rural residential location. Model 4 covers full specification consisting of labour market characteristics, human capital, demographic, sector, urban/rural residential location, and regional (province). Model 5 incorporates labour market, human capital, demographic, sector, urban/rural residential location, regional characteristics (province), and the interaction effect (ethnic-province). The table below provides information on the value of F-statistics and p-value.

Table A.1: Estimated Coefficients and F-Value: Changing the Reference

	Model 3	Model 4	Model 5	Model 3	Model 4	Model 5
Ethnic group						
Javanese	0.230**	0.056**	0.144	-0.012	0.046**	0.133+
Sundanese	0.242**	0.010	0.011	Omitted	Omitted	Omitted
Maduranese	Omitted	Omitted	Omitted	-0.242**	-0.010	-0.168
PMS	0.470**	0.076*	-0.123	0.228**	0.066*	-0.134
Betawi	0.270**	0.029	-0.021	0.028	0.019	-0.024
BMT	0.507**	0.062	0.413	0.265**	0.052**	0.363
Batak, Nias	0.491**	0.129**	0.036	0.250**	0.119**	0.026
Minang	0.474**	0.103*	0.080	0.232**	0.094+	0.070
SBS	0.509**	0.138	0.513	0.268**	0.128	0.463
Bali	0.467**	0.111	0.323	0.225**	0.109	0.312
Chinese	0.269**	-0.014	0.130	0.027	-0.023	0.120
Other ethnic groups	0.407**	0.083**	0.103	0.166**	0.073**	0.092
Province						
North Sumatera	NA	0.170**	0.042	NA	0.170**	0.213
West Sumatera	NA	0.157**	0.103	NA	0.157**	-0.069
S Sumatera & BB	NA	0.241**	0.253	NA	0.241**	0.186
Lampung	NA	0.308**	0.273	NA	0.308**	0.369**
Jakarta	NA	0.027	0.124	NA	0.027	0.154+
West Java	NA	-0.010	0.049	NA	-0.010	0.067
Central Java	NA	Omitted	Omitted	NA	Omitted	Omitted
Yogyakarta	NA	0.045*	0.003	NA	0.045*	0.334
East Java	NA	-0.245**	-0.158	NA	-0.245**	0.590*
Banten	NA	-0.033	0.089	NA	-0.033	0.082
Bali	NA	0.132*	0.415	NA	0.132*	0.142
West Nusa Tenggara	NA	0.161+	0.402	NA	0.161+	0.402
South Kalimantan	NA	0.222**	0.291	NA	0.222**	0.449
South & West Sulawesi	NA	0.250**	-0.006	NA	0.251**	0.034
Ethnicity						
F-value	71.23	2.86	0.58	71.23	2.86	0.47
p-value	<0.001	<0.001	0.85	<0.001	<0.001	0.92
Province						
F-value	NA	81.42	2.14	NA	81.42	3.89
p-value	NA	<0.001	0.01	NA	<0.001	<0.001
Interaction: Ethnic group & province						
F-value	NA	NA	15.94	NA	NA	15.94
p-value	NA	NA	<0.001	NA	NA	<0.001

Note: PMS = Palembang, Melayu, and other Sumaterans, BMT = Bugis, Makassar, Toraja, SBS = Sasak, Bima-Dompu, Sumbawa. S Sumatera & BB = South Sumatera & Bangka Belitung.

NA: Not applicable. We use $\alpha = 0.50$ for calculating earnings polarization. Model 1 uses no additional control variables (UPE). Model 2 – model 4 include additional control variables (CUPE). Model 2 captures labor market characteristics (working hours per week, occupation classification, and employment category), human capital (working experience and education level), demographic characteristics (gender, marital status, and age), and

sector (primary, secondary, or services). Model 3 covers urban/rural residential location as additional explanatory variables to Model 2. Model 4 covers regional characteristics (province) as additional explanatory variables to Model 3. Model 5 captures the interaction effect of ethnic group and province as additional explanatory variables. The primary sector includes agriculture and mining. The secondary sector includes manufacture, electricity, gas, water, and construction. Services sector includes wholesale, retail, restaurant, hotel, transportation, storage, communication, finance, insurance, real estate, and social services.

Source: The authors' calculation based on IFLS 5.

Table A.2 displays the result of sensitivity analysis for various degree of identification component (α). We conducted the calculation for $\alpha = 0.25, 0.75, \text{ and } 1$. We replicate our calculation based on Table 3. Model 1 uses no additional control variables, also known as UPE. Model 2 – model 4 include additional control variables, also known as CUPE. Model 2 captures labor market characteristics (working hours per week, occupation classification, and employment category), human capital (working experience and education level), demographic characteristics (gender, marital status, and age), and sector (primary, secondary, or services). Primary sector includes agriculture and mining. Secondary sector includes manufacture, electricity, gas, water, and construction. Services sector includes wholesale, retail, restaurant, hotel, transportation, storage, communication, finance, insurance, real estate, and social services. Model 3 covers labor market characteristics, human capital, demographic characteristics, sector, and urban/rural residential location. Model 4 covers full specification consisting of labor market characteristics, human capital, demographic characteristics, sector, urban/rural residential location, and regional characteristics (province). Model 5 adds the interaction effect of ethnic group and province to Model 4. Ethnicity and regional characteristics are significant in Model 4, where regional characteristics is stronger than ethnicity. However, ethnicity has no significant effect in Model 5, whereas regional characteristics and the interaction variable are significant.

Table A2 Estimated Coefficients and F-value in Various Degree of DER Polarization Index

	Model 1	Model 2	Model 3	Model 4	Model 5
Panel A: $\alpha = 0.25$					
Ethnic groups (estimated coefficient)					
Javanese (reference/omitted)					
Sundanese	0.113**	0.096**	0.050**	-0.075**	-0.167+
Maduranese	-0.393**	-0.384**	-0.351**	-0.135**	-0.114
Palembang, Melayu, oth Sumaterans	0.393**	0.359**	0.349**	0.011	-0.110
Betawi	0.287**	0.239**	0.119**	-0.042	-0.081
Bugis, Makassar, Toraja	0.382**	0.389**	0.384**	-0.016	0.167
Batak, Nias	0.370**	0.335**	0.373**	0.077*	0.016
Minang	0.386**	0.350**	0.348**	0.029	0.088
Sasak, Bima-Dompou, Sumbawa	0.364**	0.378**	0.363**	0.078	0.275
Bali	0.375**	0.367**	0.346**	0.072	0.279
Chinese	0.319**	0.225**	0.131+	-0.053	0.147
Other ethnic groups	0.288**	0.291**	0.242**	-0.023	0.010
Province (estimated coefficient)					
North Sumatera	NA	NA	NA	0.170**	0.167**
West Sumatera	NA	NA	NA	0.157**	0.144
S Sumatera & BB	NA	NA	NA	0.241**	0.288**
Lampung	NA	NA	NA	0.308**	0.343**
Jakarta	NA	NA	NA	0.027	-0.013
West Java	NA	NA	NA	-0.010	-0.014
Central Java	NA	NA	NA	Omitted	Omitted
Yogyakarta	NA	NA	NA	0.045*	0.042*
East Java	NA	NA	NA	-0.245**	-0.251**
Banten	NA	NA	NA	-0.033	-0.050+
Bali	NA	NA	NA	0.132*	0.016
West Nusa Tenggara	NA	NA	NA	0.161+	-0.092
South Kalimantan	NA	NA	NA	0.222**	0.224**
South & West Sulawesi	NA	NA	NA	0.250**	0.083
Ethnicity					
F-value	116.21	73.79	71.23	2.86	0.56
p-value	<0.001	<0.001	<0.001	<0.001	0.86
Province					
F-value	NA	NA	NA	81.42	72.93
p-value	NA	NA	NA	<0.001	<0.001
Interaction: Ethnic group & province					
F-value	NA	NA	NA	NA	15.94
p-value	NA	NA	NA	NA	<0.001
Panel C: $\alpha = 0.75$					
Ethnic groups (estimated coefficient)					
Javanese (reference/omitted)					
Sundanese	-0.045**	-0.045**	-0.045**	-0.032+	-0.089
Maduranese	0.059**	0.043*	0.043*	0.088**	-0.091
Palembang, Melayu, oth Sumaterans	0.010	0.002	0.002	0.044	-0.284
Betawi	-0.076**	-0.055*	-0.055*	-0.009+	0.024
Bugis, Makassar, Toraja	0.007	-0.009	-0.009	-0.005	0.092
Batak, Nias	-0.015	-0.038	-0.038	0.020	-0.274
Minang	-0.001	-0.027	-0.027	0.034	-0.173
Sasak, Bima-Dompou, Sumbawa	0.016	0.005	0.005	0.083	0.142
Bali	-0.009	-0.021	-0.021	-0.007	-0.117
Chinese	-0.084	-0.085	-0.085	-0.045	-0.190
Other ethnic groups	0.003	0.001	-0.001+	0.042+	-0.018
Province (estimated coefficient)					
North Sumatera	NA	NA	NA	-0.121**	-0.143**
West Sumatera	NA	NA	NA	-0.122*	-0.148
S Sumatera & BB	NA	NA	NA	-0.104**	-0.083*
Lampung	NA	NA	NA	-0.060*	-0.037
Jakarta	NA	NA	NA	-0.113**	-0.144**
West Java	NA	NA	NA	-0.063**	-0.044+
Central Java	NA	NA	NA	Omitted	Omitted
Yogyakarta	NA	NA	NA	-0.011	-0.013
East Java	NA	NA	NA	-0.102**	-0.109**
Banten	NA	NA	NA	-0.127**	-0.157**
Bali	NA	NA	NA	-0.070	-0.129

West Nusa Tenggara	NA	NA	NA	-0.139	-0.348
South Kalimantan	NA	NA	NA	-0.137**	-0.122*
South & West Sulawesi	NA	NA	NA	-0.053	-0.147
Ethnicity					
F-value	3.56	2.64	2.59	2.76	0.56
p-value	<0.001	<0.01	<0.01	<0.01	0.86
Province					
F-value	NA	NA	NA	6.95	7.49
p-value	NA	NA	NA	<0.001	<0.001
Interaction: Ethnic group & province					
F-value	NA	NA	NA	NA	1.81
p-value	NA	NA	NA	NA	<0.001

Panel D: $\alpha = 1.00$

Ethnic groups (estimated coefficient)

Javanese (reference/omitted)					
Sundanese	-0.102**	-0.094**	-0.077**	-0.039+	-0.092
Maduranese	0.228**	0.206**	0.194**	0.172	-0.126
Palembang, Melayu, oth Sumaterans	-0.129**	-0.128**	-0.125**	0.062+	-0.262
Betawi	-0.183**	-0.140**	-0.095**	-0.007	0.034
Bugis, Makassar, Toraja	-0.149**	-0.171**	-0.169**	-0.022	0.038
Batak, Nias	-0.172**	-0.197**	-0.211**	-0.011	-0.381+
Minang	-0.152**	-0.182**	-0.182**	0.025	-0.212
Sasak, Bima-Dompu, Sumbawa	-0.140**	-0.159**	-0.153**	0.115	0.053
Bali	-0.154**	-0.167**	-0.160**	-0.041	-0.339
Chinese	-0.174*	-0.169*	-0.135	-0.017	-0.263
Other ethnic groups	-0.117**	-0.116**	-0.098**	0.014	0.047

Province (estimated coefficient)

North Sumatera	NA	NA	NA	-0.294**	-0.329**
West Sumatera	NA	NA	NA	-0.284**	-0.339+
S Sumatera & BB	NA	NA	NA	-0.305**	-0.297**
Lampung	NA	NA	NA	-0.272**	-0.253**
Jakarta	NA	NA	NA	-0.174**	-0.201**
West Java	NA	NA	NA	-0.081**	-0.047
Central Java	NA	NA	NA	Omitted	Omitted
Yogyakarta	NA	NA	NA	-0.059*	-0.059*
East Java	NA	NA	NA	-0.032+	-0.040*
Banten	NA	NA	NA	-0.163**	-0.202**
Bali	NA	NA	NA	-0.180*	-0.212
West Nusa Tenggara	NA	NA	NA	-0.348**	-0.489
South Kalimantan	NA	NA	NA	-0.333**	-0.305**
South & West Sulawesi	NA	NA	NA	-0.215**	-0.274

Ethnicity

F-value	21.89	20.98	18.52	4.12	0.53
p-value	<0.001	<0.001	<0.001	<0.001	0.88

Province

F-value	NA	NA	NA	15.80	11.98
p-value	NA	NA	NA	<0.001	<0.001

Interaction: Ethnic group & province

F-value	NA	NA	NA	NA	4.00
p-value	NA	NA	NA	NA	<0.001

Note: Model 1 uses no additional control variables (UPE). Model 2 – model 4 include additional control variables (CUPE). Model 2 captures labor market characteristics (working hours per week, occupation classification, and employment category), human capital (working experience and education level), demographic characteristics (gender, marital status, and age), and sector (primary, secondary, or services). Model 3 covers urban/rural residential location as additional explanatory variables to Model 2. Model 4 covers regional characteristics (province) as additional explanatory variables to Model 3. Model 5 captures the interaction effect of ethnic group and province as additional explanatory variables. The primary sector includes agriculture and mining. The secondary sector includes manufacture, electricity, gas, water, and construction. Services sector includes wholesale, retail, restaurant, hotel, transportation, storage, communication, finance, insurance, real estate, and social services.

Source: The author's calculation based on IFLS 5.

Table A.3 exhibits the result of the sensitivity analysis in which we re-classify the ethnic groups. We re-classify all ethnic groups into seven groups (Panel A), three groups (Panel B), and two groups (Panel A). Model 1 uses no additional control variables. Model 2 captures labour market characteristics (working hours per week, occupation classification, and employment category), human capital (working experience and education level), demographic characteristics (gender, marital status, and age), and sector (primary, secondary, or services). Model 3 covers labour market characteristics, human capital, demographic characteristics, sector, and urban/rural residential location. Model 4 covers a full specification variable consisting of labour market characteristics, human capital, demographic characteristics, sector, urban/rural residential location, and regional characteristics (province). Model 5 provides additional complexity by incorporating the interaction variables between ethnicity and province. In Model 4, ethnicity and regional characteristics are significant, but regional characteristics are stronger than ethnicity. In Model 5, ethnicity has no significant effect anymore, whereas regional characteristics and the interaction variable are significant.

Table A3 Estimated coefficients of RIF-regression (Re-classification of ethnic groups)

Coefficients	Model 1	Model 2	Model 3	Model 4	Model 5
Panel A: 7 ethnic groups					
Ethnic groups (estimated coefficient)					
Javanese (reference/omitted)					
Sundanese & Betawi	0.075**	0.060**	0.017	-0.042**	-0.134 ⁺
Maduranese	-0.248**	-0.253**	-0.231**	-0.056**	-0.076
Batak, Nias, Minang, PMS	0.273**	0.244**	0.248**	0.041*	-0.105
Bali, Sasak, Bima-Dompu, Sumbawa	0.275**	0.272**	0.259**	0.063	0.180
Bugis, Makassar, Toraja	0.285**	0.281**	0.277**	0.004	0.180
Other ethnic groups	0.210**	0.204**	0.167**	0.017	-0.019
Province (estimated coefficient)					
North Sumatera	NA	NA	NA	0.184**	0.166**
West Sumatera	NA	NA	NA	0.163**	0.145
S Sumatera & BB	NA	NA	NA	0.229**	0.287**
Lampung	NA	NA	NA	0.305**	0.342**
Jakarta	NA	NA	NA	0.030	-0.012
West Java	NA	NA	NA	-0.011	-0.13
Central Java	NA	NA	NA	Omitted	Omitted
Yogyakarta	NA	NA	NA	0.046*	0.044*
East Java	NA	NA	NA	-0.248**	-0.251**
Banten	NA	NA	NA	-0.033	-0.050 ⁺
Bali	NA	NA	NA	0.124*	0.018
West Nusa Tenggara	NA	NA	NA	0.179**	-0.090
South Kalimantan	NA	NA	NA	0.229**	0.223**
South & West Sulawesi	NA	NA	NA	0.251**	0.084
Ethnicity					
F-value	143.75	132.64	129.20	4.38	0.96
p-value	<0.001	<0.001	<0.001	<0.001	0.45
Province					
F-value	NA	NA	NA	81.69	72.72
p-value	NA	NA	NA	<0.001	<0.001
Interaction: Ethnic group & province					
F-value	NA	NA	NA	NA	23.54
p-value	NA	NA	NA	NA	<0.001
Panel B: 3 ethnic groups					
Ethnic groups (estimated coefficient)					
Javanese (reference/omitted)					
Sundanese & Betawi	0.075**	0.060**	0.017	-0.049**	-0.134 ⁺
Other ethnic groups	0.181**	0.167**	0.162**	-0.002	-0.060
Province (estimated coefficient)					
North Sumatera	NA	NA	NA	0.206**	0.167**
West Sumatera	NA	NA	NA	0.197**	0.147
S Sumatera & BB	NA	NA	NA	0.251**	0.286**
Lampung	NA	NA	NA	0.312**	0.341**
Jakarta	NA	NA	NA	0.035 ⁺	-0.010
West Java	NA	NA	NA	-0.005	-0.012

Central Java	NA	NA	NA	Omitted	Omitted
Yogyakarta	NA	NA	NA	0.045*	0.044*
East Java	NA	NA	NA	-0.255**	-0.250**
Banten	NA	NA	NA	-0.026	-0.050+
Bali	NA	NA	NA	0.178**	0.019
West Nusa Tenggara	NA	NA	NA	0.239**	-0.093
South Kalimantan	NA	NA	NA	0.239**	0.222**
South & West Sulawesi	NA	NA	NA	0.254**	0.084
Ethnicity					
F-value	165.19	141.30	145.32	6.60	1.80
p-value	<0.001	<0.001	<0.001	<0.01	0.17
Province					
F-value	NA	NA	NA	119.92	72.34
p-value	NA	NA	NA	<0.001	<0.001
Interaction: Ethnic group & province					
F-value	NA	NA	NA	NA	47.93
p-value	NA	NA	NA	NA	<0.001

Panel C: 2 ethnic groups

Ethnic groups (estimated coefficient)

Javanese+ (reference/omitted)					
Other ethnic groups	0.246**	0.235**	0.234**	0.044**	-0.077
Province (estimated coefficient)					
North Sumatera	NA	NA	NA	0.181**	0.167**
West Sumatera	NA	NA	NA	0.158**	0.112
S Sumatera & BB	NA	NA	NA	0.226**	0.273**
Lampung	NA	NA	NA	0.296**	0.316**
Jakarta	NA	NA	NA	0.005	0.005
West Java	NA	NA	NA	-0.041**	-0.049**
Central Java	NA	NA	NA	0.046*	0.046*
Yogyakarta	NA	NA	NA	Omitted	Omitted
East Java	NA	NA	NA	-0.254**	-0.257**
Banten	NA	NA	NA	-0.053**	-0.052*
Bali	NA	NA	NA	0.177**	0.184**
West Nusa Tenggara	NA	NA	NA	0.199**	0.080
South Kalimantan	NA	NA	NA	0.207**	0.222**
South & West Sulawesi	NA	NA	NA	0.214**	0.084
Ethnicity					
F-value	561.33	510.76	536.52	9.38	0.52
p-value	<0.001	<0.001	<0.001	<0.01	0.47
Province					
F-value	NA	NA	NA	99.35	92.94
p-value	NA	NA	NA	<0.001	<0.001
Interaction: Ethnic group & province					
F-value	NA	NA	NA	NA	70.36
p-value	NA	NA	NA	NA	<0.001

Note: PMS = Palembang, Melayu, & other Sumaterans. Javanese+ = Javanese, Sundanese, Banten, Cirebon, Betawi, Maduranese, Bali. The reference group is Javanese in Panel A and Panel B, and Javanese+ in Panel C.

Model 1 uses no additional control variables (UPE). Model 2 – model 4 include additional control variables (CUPE). Model 2 captures labor market characteristics (working hours per week, occupation classification, and employment category), human capital (working experience and education level), demographic characteristics (gender, marital status, and age), and sector (primary, secondary, or services). Model 3 covers urban/rural residential location as additional explanatory variables to Model 2. Model 4 covers regional characteristics (province) as additional explanatory variables to Model 3. Model 5 captures the interaction effect of ethnic group and province as additional explanatory variables. The primary sector includes agriculture and mining. The secondary sector includes manufacture, electricity, gas, water, and construction. Services sector includes wholesale, retail, restaurant, hotel, transportation, storage, communication, finance, insurance, real estate, and social services.

Source: The authors' calculation based on IFLS 5.