WIDER Working Paper 2022/12

Whose intergenerational mobility?

A new set of estimates for Indonesia by gender, geography, and generation

Diding Sakri,[1] Andy Sumner,[2] and Arief Anshory Yusuf[3]

February 2022

Abstract: Various scholars have estimated levels of intergenerational mobility in OECD countries. Fewer estimates are available for developing countries, where mobility arguably matters more due to starker differences in living standards. This paper presents new estimates of mobility for a developing country, namely Indonesia. The estimates are based on data from five waves of the Indonesian Family Life Survey, a longitudinal analysis of socio-economic status which began in 1993. We constructed a pooling sample consisting of 9,445 matching pairs of children and their parents. The paper estimates relative mobility using the intergenerational elasticity of income. We find that although, overall, the intergenerational elasticity of income is low compared with other developing countries, the level of mobility in Indonesia differs markedly by children's gender and across provinces and generations.

Key words: intergenerational mobility, longitudinal analysis, Indonesia, Indonesian Family Life Survey

JEL classification: D63, J62

Acknowledgements: We thank Kunal Sen and Armando Barrientos for comments on the PhD thesis on which this paper is based and Eduardo Ortiz-Juarez for comments on the draft paper. The lead author thanks the Indonesia Endowment Fund for Education (LPDP) for funding his PhD Research, from which this paper derives.

[1] West Java Governor's Advisory Team, Indonesia, corresponding author: andrew.sumner@kcl.ac.uk; [2] King's College London, UK; [3] Padjadjaran University, Indonesia

This study has been prepared within the UNU-WIDER project Social mobility in the Global South – concepts, measures, and determinants.

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9267-143-3

https://doi.org/10.35188/UNU-WIDER/2022/143-3

Typescript prepared by Joseph Laredo.

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

1 Introduction

Various scholars have estimated levels of intergenerational mobility in OECD countries. The available empirical evidence on relative mobility most often focuses on international comparison of estimates and the 'Great Gatsby Curve' (Corak 2013; DiPrete 2020; Durlauf et al. 2021; Iversen et al. 2021; Krueger 2012; Narayan et al. 2018; OECD 2008, 2011), which depicts the relationship between income inequality and the intergenerational elasticity of income.[1]

In developing countries, however, fewer mobility estimates are available, although mobility matters more in terms of the ability to progress to higher standards of living above an absolute poverty line (Iversen et al. 2021). A major constraint—among many—on the analysis of intergenerational mobility in developing countries is the absence of representative longitudinal data. Unlike (repeated) cross-sectional analysis, which provides anonymous information about the income share of segments of society (and changes in their share over time), longitudinal analysis reveals changes occurring in the same household. If longitudinal data are collected for sufficient time to account for at least two generations (parents and their children), it is possible to estimate various measures of economic mobility across generations. The Indonesian Family Life Survey (IFLS) has collected such longitudinal data for Indonesia from 1993 (first wave) up to 2014 (fifth wave). We use this dataset in the current paper.

The same data source, IFLS, was used by Dartanto et al. (2019) to study *intragenerational* rather than *intergenerational* economic mobility between 1993 and 2014. They define mobility as moving in and out of poverty as well as transitioning from poverty into the middle or upper class, with the household instead of the individual as the unit of analysis. In contrast to Dartanto et al. (2019), we construct a pooling sample with 9,445 matching pairs of children and their parents (thus individuals, not households) and compare their economic outcomes in terms of level and position in the income distribution (intergenerational mobility) rather than changes in parents' income over time (intragenerational mobility). Subsequently, two measures of relative mobility are estimated, namely intergenerational elasticity based on log-log regression of children's outcomes on their parents' outcomes and rank-rank specification. Furthermore, we consider how mobility differs by gender (of the child), by birth cohort (pre-millennials, born before 1980, versus millennials, born in or after 1980), and by provincial origin of the parents. We find that although, overall, the intergenerational elasticity of income is low compared with other developing countries, the level of mobility in Indonesia differs markedly by children's gender and across provinces and generations.

This paper is structured as follows: Section 2 discusses the concepts and measurements of mobility. Section 3 details our methodology. Section 4 discusses our estimates of mobility in Indonesia. Section 5 then considers nonlinearity and mobility by children's gender, across Indonesia's

[1] The 'Great Gatsby Curve' was first highlighted by Judd Cramer, a staff economist of the US Council of Economic Advisors and then by the Chairman of the Council of Economic Advisers, Alan Krueger (2012). The curve shows that the association between parental economic status and adult earnings of children is weakest in countries with low income inequality, such as Finland, Norway, Denmark, and Sweden: less than one-fifth of any economic advantage or disadvantage that a parent may have experienced in his time is passed on to a child in adulthood (Corak 2013). Conversely, in countries with higher inequality, like Italy, the United Kingdom, and the United States, roughly 50 per cent of any advantage or disadvantage is handed down (Corak 2013). This evidence suggests that in relatively unequal societies, children of parents with relatively low income are much more likely to ultimately hold relatively low-income positions themselves. From a meritocratic point of view this is problematic, since it signals that economic outcomes are associated with parental background rather than with individual effort and talent.

provinces and between pre-millennials and millennials. Section 6 concludes. An annex describes each step of our methodology in detail.

2 Concepts and measures of intergenerational mobility

2.1 Mobility as a multifaceted concept

Mobility is a multifaceted and multidimensional concept (see discussion in Iversen et al. 2021). Therefore, studies are often not comparable conceptually nor methodologically, as they use different concepts and measurements of mobility. The seminal studies of Sorokin (1927) and Glass (1954) on the US and the UK, respectively, are often cited as the earliest work in the field. And while some studies have made estimates of inter- and intra-generational mobility, others have focused on absolute and relative mobility or compared educational, occupational, and income mobility. Solon's (1999) literature survey on mobility elaborated intergenerational elasticity (IGE, β coefficient) as a measure of mobility within a framework of labour earnings and as a return on the investment of parents in the human capital of their children. More than a decade later, Black and Devereux (2011) updated this survey to include numerous empirical studies applying the framework introduced in Solon's paper. However, as Fields and Ok (1999: 557) noted, there is no unified framework for mobility because even the notion of income mobility has multiple definitions; thus different studies adopt different approaches. More recent surveys include those of Jäntti and Jenkins (2015), Iversen et al. (2021), and Deutscher and Mazumder (forthcoming), each elaborating frameworks and highlighting key concepts, measures, and properties.

Fields (2005: 7–14; see also Fields 2021) proposed six different notions of mobility that apply to both intragenerational changes (within the same generation over time) and intergenerational studies (between different generations). Following this, intergenerational mobility can refer to: (i) origin dependence, which considers the extent to which parents' economic well-being determines that of their children; (ii) positional movement, which compares children's economic position among their peers (ranks, centiles, deciles, or quintiles) with the economic position of their parents relative to the latter's peers; (iii) share movement, which analyses how children's shares of the total income of their generation differ from the shares of their parents relative to their respective generation; (iv) income flux, which investigates the extent of fluctuation between parents' incomes and the incomes of their children but not the direction of the change; (v) directional income movement, which is concerned with the number of parents–children pairs that move up or down and by how much; and (vi) mobility as an equalizer of longer-term incomes, which compares the income inequality within the parental generation with the inequality within the children's generation.[2] In short, there are numerous concepts and measures of mobility, raising problems for comparability (Fields and Ok 1999; Fields 2021; Iversen et al. 2021; Jäntti and Jenkins 2015).

In a developing country context, intergenerational mobility studies are particularly challenging because there are fewer longitudinal datasets and because of the difficulties in estimating income where agrarian and informal employment are widespread.[3]

[2] See also Savegnago (2016), who summarizes various indices of mobility including their formulas and references.

[3] Narayan et al.'s (2018) approach is to use retrospective data on parental education in developing countries as a measure of intergenerational mobility.

2.2 Concepts and measures of relative mobility

In this paper we focus on relative mobility and income. Relative mobility measures the degree to which the economic ranking of adult children among their peers is independent of their parents' ranking relative to their respective peers (Narayan et al. 2018: 52). Relative mobility can be interpreted as origin independence (i.e. persistence), meaning that the personal characteristics of children (such as talent or education level) rather than their parental background (e.g. occupation, social status, or income position) determine economic outcomes (in keeping with Roemer 2004). This is based on the meritocratic idea that an individual's life chances should depend on her own abilities and effort rather than on who her parents are.

The canonical measure of relative mobility is intergenerational elasticity (IGE). IGE is usually derived as the least-squares estimate of the coefficient β in the following equation:

$$\log\left(\overline{y}_{i,g1}\right) = \alpha + \beta \log\left(\overline{y}_{i,g0}\right) + \epsilon_{i,g1} \tag{1}$$

where $\overline{y}_{i,g1}$ and $\overline{y}_{i,g0}$ represent the mean economic outcome of the children's and the parents' generation, respectively. Accordingly, $\epsilon_{i,g1}$ denotes all other influences on adult children's outcomes not correlated with parents' outcomes. The constant term α captures the trend in average outcome across generations due to, for example, changes in productivity, international trade, technology, or labour market institutions. The equation was introduced in the context of mobility by Becker and Tomes (1986: 2) and is the standard economic model of intergenerational mobility (see discussion in Piraino 2021).

IGE indicates the degree to which outcomes are 'sticky' across generations of the same family by estimating the percentage difference in children's outcomes for each percentage point difference in parents' outcomes. It represents the fraction of economic advantage that is on average transmitted across generations. In other words, β summarizes in a single number the degree of intergenerational income mobility in a society. A positive value indicates intergenerational persistence of incomes in the sense that higher parental incomes are associated with higher incomes of children. In turn, a negative value implies reversal of incomes, manifested in higher parental incomes correlated with lower incomes of children.

Empirical studies in OECD countries have found β to always lie between zero and one. The higher the value of β, the higher the predictability of children's future economic ranking based on the observable position of their parents in the income distribution. The lower the value of β, the less 'stickiness'. In other words, when β is low, then parents' relative outcomes are a weak predictor of their children's future position in the income distribution of their own generation. Hypothetically, following Corak (2013), β=0 represents a case of complete mobility where the outcomes of parents and children are entirely unrelated, while β=1 represents a case of complete immobility with the proportionate (dis)advantage of parents being mirrored one-to-one in their children's generation.

In the equation above, both economic outcomes are presented in logarithmic terms. However, β estimations resulting from this log-on-log equation have two important limitations. First, the relationship between log child income and log parent income is highly nonlinear. This was not apparent in earlier empirical works due to smaller samples. As a result of this nonlinearity, IGE is sensitive to the point of measurement in the income distribution, as shown in many studies (for instance, Björklund et al. 2012; Bratsberg et al. 2007; Chetty et al. 2014a; Chetty et al. 2014b; Corak et al. 2014; Grawe and Mulligan 2002; Gregg et al. 2019). Second, the log-log specification discards observations with zero income, which often account for a substantial fraction of the sample. Dropping zero income observations might therefore overstate the degree of intergenerational mobility. In other words, the way in which these zero income observations are treated can change the IGE estimate dramatically; see for instance Chetty et al. (2014b) and Gregg et al. (2019).

An alternative measure of relative intergenerational mobility which considers these limitations is the correlation between child rank and parent rank. This has been found to be almost perfectly linear and highly robust to alternative specifications (e.g. in Bratberg et al. 2017; Chetty et al. 2014b; Corak et al. 2014; Nybom and Stuhler 2015; Pekkarinen et al. 2017). Thus, the equation above can be modified to:

$$\overline{PRy}_{i,g1} = \alpha + \rho\left(\overline{PRy}_{i,g0}\right) + \epsilon_{i,g1} \quad (2)$$

Let $\overline{PRy}_{i,g1}$ denote child i's percentile rank in the income distribution of children (generation 1) and $\overline{PRy}_{i,g0}$ represent the parents i's percentile rank in the income distribution of parents (generation 0). Importantly, this definition allows us to include observations with zero income in generation 1. Regressing the child's rank $\overline{PRy}_{i,g1}$ on her parents' rank $\overline{PRy}_{i,g0}$ yields a regression coefficient ρ which equals $Corr(\overline{PRy}_{i,g1}, \overline{PRy}_{i,g0})$.

3 Methodology

3.1 How previous studies constructed the core sample of mobility research

To conduct intergenerational mobility research, one has to construct core sample of pairs of parents and their children (two-generational copulas) or pairs across more than two generations. This first step is to analyse the relationship between the economic performance of parents and that of their children. However, the period of productivity of adult children is different from the productivity period of their parents. Consequently, constructing such a sample is challenging since it must contain economic outcomes data for both generations. Previous empirical studies have used different data sources and methods to create such a sample. For example, Atkinson et al. (1978) traced children of families who participated in a survey carried out 26 years earlier. Out of 2,011 original respondents surveyed in 1950, only 1,363 could be traced by address and name and subsequently mailed. This group of 1,363 traceable families was composed of 260 families who ultimately were not trackable, 57 who refused to answer, and 220 who did not have children and hence could not be part of the mobility sample, leaving 826 families with a total of 2,236 children who could be included in the later stage of the survey to collect income data for adult children. Of these 2,236 children, only 1,348 from a total of 500 families could be located for income survey interviews. In short, after various follow-up surveys, the final sample eligible for mobility analysis consisted of 307 father–son pairs. As reported in Atkinson (1980), this great reduction is due to several factors such as excluding those fathers from the sample who did not work when surveyed in 1950 or equally those children not employed when surveyed in 1975–78.

The method applied by Atkinson et al. (1978) is nevertheless arguably better than a more straightforward method where both parents' and children's incomes are measured at the same time. The principal shortcoming of the latter method, reported in Atkinson (1980), is that fathers and children are at very different stages of their life cycles at the point of comparison. Many of the children might not yet be independent of their parents, and even if parents are selected from older age groups, a sizeable number of children might have entered the labour force only relatively recently, which implies lower incomes in many cases.

A more desirable method than either of the above-mentioned approaches is a longitudinal survey that encompasses responses from at least two different points in time that are sufficiently far apart to record income data of both parents and their adult children. An oft-cited research of this kind is that of Solon (1992), who obtained data from the Panel Study of Income Dynamics (PSID). This is a nationally representative longitudinal survey of about 5,000 families in the United States that the University of Michigan's Survey Research Center has conducted annually since 1968. Solon's study focused mainly on father–son correlations in earnings, hourly wage rates, and family income, with a main sample comprising 348 father–son pairs extracted from the PSID. The sons in the sample were children from the original 1968 PSID households who, in the 1985 survey, reported positive annual earnings for 1984. The sons' sample was restricted to the cohort born between 1951 and 1959: sons born before 1951 were excluded to avoid over-representation of sons who left home at a late age, while the 1959 restriction ensured that the measurement of sons' statuses in 1984 were observed at ages of at least 25 (Solon 1992: 397).

Another approach employs multiple administrative data to establish links between the children's and parents' generation such as in Chetty et al. (2014a) and Chetty et al. (2014b). The authors of both these papers constructed a linked parents–children sample using population tax records from 1996 to 2012 encompassing all individuals born between 1980 and 1993 who were US citizens as of 2013 and were indicated as a dependant on a tax return filed in or after 1996. The researchers linked approximately 95 per cent of children in each birth cohort to their parents based on dependant-claiming, obtaining a core sample with 3.7 million children per cohort and 40 million children in total. Although this approach is undoubtedly powerful in serving empirical research, it is difficult to replicate in other contexts, particularly in countries where such tax and administrative records are not easily accessible for research purposes, which is frequently the case in developing countries.

3.2 Constructing potential core samples for mobility analysis in Indonesia

In the context of Indonesia, one survey permits the construction of a core sample for mobility research, namely the Indonesian Family Life Survey (IFLS). The IFLS is an ongoing longitudinal socioeconomic and health survey based on a sample of households that is representative of about 83 per cent of the Indonesian population and contains individuals living in 13 of the nation's 26 provinces. The first wave (IFLS1) was executed in 1993 with individuals living in 7,224 households, followed by IFLS2 (1997), IFLS3 (2000), and IFLS4 (2007), which tracked the original households from 1993 and their split-offs, which by IFLS5 (2014) totalled 16,204 households and 50,148 individuals interviewed (for more details see Strauss et al. 2016). The original households of 1993 that had one child or more and could be tracked until the latest survey in 2014 were labelled as dynastic households. Dynastic households account for 80.2 per cent (5,794) of the original households and are distributed across all 13 provinces considered (see Table 1).

Table 1: Provincial distribution of sample: original vs dynastic households

(1) Province	(2) Original HH (n)	(3) Original HH (%)	(4) Dynastic HH (n)	(5) Dynastic HH (%)	(6) Dynastic/ original HH (%)
Bali	340	4.71	287	4.95	84.41
Yogyakarta	478	6.62	322	5.56	67.36
Jakarta	731	10.12	590	10.18	80.71
Jabar	1,111	15.38	872	15.05	78.49
Jateng	878	12.15	702	12.12	79.95
Jatim	1,044	14.45	796	13.74	76.25
Kalsel	323	4.47	262	4.52	81.11
Lampung	274	3.79	238	4.11	86.86
NTB	407	5.63	341	5.89	83.78
Sulsel	375	5.19	323	5.57	86.13
Sumbar	351	4.86	277	4.78	78.92
Sumsel	349	4.83	299	5.16	85.67
Sumut	563	7.79	485	8.37	86.15
Total	*7,224*	*100.00*	*5794*	*100.00*	*80.20*

Source: authors' estimates based on IFLS.

Dynastic households (5,794) are the source from which children–parents copulas can be constructed. The number of IFLS respondents and the relatively long time span of the survey (21 years between the first and the latest) enables the construction of a substantial core sample for intergenerational mobility analysis compared with that used in similar studies (see Narayan et al. 2018). It is possible to use all waves of the IFLS to construct nine pairs: (1) fathers–sons, (2) fathers–daughters, (3) fathers–children, (4) mothers–sons, (5) mothers–daughters, (6) mothers–children, (7) parents–sons, (8) parents–daughters, and (9) parents–children. The complete list of copulas derived from the IFLS is presented in Table 2. Note that the total number of dynastic households (5,794), which is indicated in the final row in the second column of Table 2, corresponds to the value recorded in the final row of column 4 of Table 1.

Table 2: All copulas of parents and their children derived from IFLS: potential core samples for mobility study

(1) Copula	(2) Dynasty (number of dynastic HH)	(3) Children (number)	(4) Mean age of child[a]	(5) Mean age of parent[b]	(6) Mean years of schooling, child	(7) Mean years of schooling, parent[c]
(1) Fathers–sons	3,949	6,937	32.67	44.37	10.70	7.17
(2) Fathers–daughters	3,818	6,433	32.53	44.02	10.73	7.24
(3) Fathers–children	5,119	13,370	32.60	44.20	10.71	7.20
(4) Mothers–sons	4,296	7,516	33.15	39.18	10.63	5.91
(5) Mothers–daughters	4,211	7,003	33.10	39.08	10.57	5.88
(6) Mothers–children	5,669	14,519	33.13	39.13	10.60	5.90
(7) Parents–sons	4,396	7,675	33.26	42.05	10.62	7.50
(8) Parents–daughters	4,307	7,161	33.23	41.91	10.59	7.47
(9) Parents–children	5,794	14,836	33.25	41.98	10.61	7.48

Note: [a] measured in 2014; [b] measured in 1993, age of older parent considered in copulas 7 to 9; [c] higher value among both parents considered in copulas 7 to 9.

Source: authors' estimates based on IFLS.

Table 2 shows the richness of the IFLS data in the sense of how large the dataset is relative to the usual parent–children data used in the literature. In the IFLS, 5,794 households with 14,836 children represent our potential core sample for mobility analysis. However, this potential core sample represents only cases where a link has been established between parents and their children; it does not consider the availability of earnings data for parents and their children, which are available for only 8,889 cases. We do *not* impute missing income data. Thus we use 8,889 pairs of children and parents, which is still a large sample compared with the typical sample of a few hundred observations in the literature (see Narayan et al. 2018) and remains representative of Indonesia.[4]

3.3 From potential to actual core samples: availability of earnings data

Table 2 shows that the longitudinal nature of IFLS data makes it possible to construct all nine potential copulas of mobility analysis. In this paper, the analysis will first focus on the parent–children copula type (type 9 or row 9 in Table 2). As shown in Table 2, parents–children copulas encompass 14,836 children from 5,794 parents, which means we have a potential of 14,836 children whose economic outcome data may be compared with their respective parents' income data. The number of parents–children copulas actually eligible for intergenerational income mobility analysis depends on the availability of data to measure our main variable, which is economic outcome, for both parents and children.

In this paper, personal income is used as economic outcome. There are three steps to estimate the latter in this research. First, we define covariates of the income variable to be used in the estimation. There are five covariates: working time, occupation, employment type, sector, and geographical location of the respondent's workplace. Second, we extract earnings data as reported by respondents. The last step of our earnings estimation applies temporal and spatial deflators to estimate real values in addition to the nominal values reported in the IFLS. Details of the three steps in our earnings data estimation are described in the Appendix.

3.4 Issues concerning the quality of earnings data

There are three issues concerning the quality of the data, namely coresidency bias, lifecycle bias, and transitory income shocks. These issues should be considered carefully when creating the actual core sample for intergenerational income mobility analysis. We deal with these issues as follows.

Intergenerational mobility samples (matching parents and their children) constructed from household surveys are sensitive to sample bias (coresidency bias). This is because the IFLS and other standard household surveys, such as the Living Standards Measurement Survey (LSMS) of the World Bank, usually include *only* coresident parents and children; they do not gather any information on family members who do not satisfy the coresidency criterion. Thus, according to Emran et al. (2016), coresidency restrictions result in a truncated sample. Since the pattern of coresidency is not random, most of the studies suffer from potentially serious bias when estimating intergenerational persistence in economic status.

One way to check for coresidency bias is to compare mobility estimates derived from standard household surveys, as used in this paper, with those from another sample of the same population (in this case Indonesia) that does not apply coresidency restrictions in the sampling process. However, to the best of the authors' knowledge, there are no such data accessible for Indonesia. Fortunately, the IFLS tracked all respondents longitudinally in all surveys after its first wave in 1993 and updated the status of respondents' residence during each new wave. Hence, for each adult child we were able to identify the household (s)he currently belongs to as well as her/his

[4] See Sakri (2019: 70–81, 84–89) for full details.

original household. In effect, we were able not only to create a pooling sample containing all the children analysed but also to use coresidency status as a control in our pooling sample.

In addition to coresidency bias, mobility estimates are sensitive to the age of both parents and children at the point in time when incomes are measured and therefore subject to lifecycle bias. Nybom and Stuhler (2016) show that IGE estimates can vary substantially with the age at which sons' (children in our case) incomes are observed, and that the bias is smallest when incomes are observed around midlife. When incomes are not observed around midlife, other researchers such as Deng et al. (2013) and Ferreira and Veloso (2006) have tried to minimize the bias by including the age of the child and parent as well as their age squared in the estimating equation and subsequently comparing the results with the baseline estimates. We followed this approach in our present study. Additionally, we included a dummy variable for millennial children as a control. We define children as millennials if they were born in or after year 1980.

Finally, mobility estimates are sensitive to transitory income shocks, which suggests that averaging several observations of income at different points in time will produce results closer to the actual lifetime income than those derived from a one-time income observation. Therefore, in this paper, we average earnings data, after standardizing them with spatial and temporal deflators, from up to five observation points if data availability allows it. The number of observation points used to average income is further outlined in the summary statistics of the data below.

4 Intergenerational mobility in Indonesia

4.1 Summary statistics of the intergenerational mobility core sample

The core sample resulting from the data construction process described in Section 3 and the Appendix is summarized in Table 3. We find that earnings data for children are recorded on average twice out of the possible five times compared with thrice for parents. This implies that the earnings data used for our intergenerational analysis are typically averaged from more than one observation for both children and parents, thus reducing the risk of bias due to transitory income shocks, as outlined in the previous discussion. Furthermore, Table 3 illustrates that children were on average 34 years old in 2014 when their incomes were recorded, compared with their parents' average age of 43 years in 1993 when their earnings were registered. This indicates that both average ages fall within the productive age range, although there is a sizeable difference between them. This leads us to include age as a control variable when estimating the mobility index.

Table 3: Summary statistics of intergenerational income mobility core sample in Indonesia: parents–children

	Pooling sample					
Statistic	**N**	**Min.**	**Median**	**Mean**	**SD**	**Max.**
Log earning, child	9,445	7.97	13.22	13.14	1.16	16.74
Log earning, parent	8,889	7.29	12.63	12.59	1.11	16.22
Rounds of data, child	9,445	1	2	1.85	0.97	5
Rounds of data, parent	9,445	0	3	2.80	1.28	5
Age, child	9,445	21	33	34.25	8.12	80
Age, parent	9,445	18.5	41.5	42.99	10.61	88
Millennials	9,445	0	1	0.56	0.50	1
Male	9,445	0	1	0.58	0.49	1
Coresident=1						
Log earning, child	4,194	7.97	13.04	12.97	1.18	16.73
Log earning, parent	3,856	7.78	12.64	12.57	1.16	16.22
Rounds of data, child	4,194	1	1	1.66	0.89	5
Rounds of data, parent	4,194	0	3	2.67	1.34	5
Age, child	4,194	21	33	34.29	9.09	80
Age, parent	4,194	19.5	42	43.68	11.33	88
Millennials	4,194	0	1	0.57	0.50	1
Male	4,194	0	1	0.59	0.49	1
Coresident=0						
Log earning, child	5,250	8.21	13.36	13.27	1.12	16.74
Log earning, parent	5,033	7.29	12.63	12.62	1.06	16.22
Rounds of data, child	5,250	1	2	2.00	1.00	5
Rounds of data, parent	5,250	0	3	2.90	1.21	5
Age, child	5,250	21	34	34.22	7.25	73
Age, parent	5,250	18.5	41	42.44	9.95	86
Millennials	5,250	0	1	0.55	0.50	1
Male	5,250	0	1	0.57	0.50	1

Source: authors' estimates based on IFLS.

4.2 Intergenerational mobility in Indonesia

This section reports estimates of mobility in Indonesia derived from the pooling sample. The mobility estimates based on equation 2 were found to be robust to coresidency bias, lifecycle bias, and transitory income shocks and can therefore be used as reference estimates of mobility in Indonesia. Table 4 illustrates β estimates for the three samples, ranging from 0.291 (coresident=0) to 0.326 (coresident=1), signifying that the non-coresident sample demonstrates higher mobility.

Table 4: Comparison of IGE estimates based on log-log specification

β	**Non-coresident**	**Pooling**	**Coresident=1**
EarningG0	0.291***	0.310***	0.326***
SE	(0.015)	(0.011)	(0.016)
Lower-bound	0.261	0.289	0.295
Upper-bound	0.321	0.332	0.357
Constant	9.611***	9.245***	8.891***
SE	(0.193)	(0.139)	(0.198)
R-squared	0.075	0.088	0.103
N	5,033	8,889	3,856

Note: * p<0.05, ** p<0.01, *** p<0.001

Source: authors' estimates based on IFLS.

The difference in β estimates between the coresident and non-coresident samples suggests that children with higher earnings are potentially the children moving out of their parents' household and living on their own. This assumption is supported by the fact that the median and mean earnings of those who have moved out (coresident=0) are US$50.98 and US$82.13, respectively, which are higher than the median and mean earnings of coresident children, US$37.12 and US$65.10.

IGE was also estimated by rank-rank specification (ρ) as in equation 2. ρ estimates (reported in Table 5) range from 0.277 (coresident=0) to 0.313 (coresident=1) and thus are found to be systematically higher than β estimates (illustrated in Table 4). Like the estimates of β, the ρ estimates of the non-coresident sample demonstrate higher mobility than those of the counterpart sample. This strengthens the above-mentioned hypothesis that it is children with higher earnings who establish their own independent households.

Table 5: Comparison of IGE estimates based on rank-rank specification

ρ	Coresident=0	Pooling	Coresident=1
EarningG0	0.277***	0.299***	0.313***
SE	(0.014)	(0.010)	(0.014)
Lower-bound	0.250	0.280	0.286
Upper-bound	0.303	0.319	0.340
Constant	39.665***	35.419***	30.891***
SE	(0.795)	(0.563)	(0.781)
R-squared	0.074	0.091	0.109
N	5,251	9,445	4,194

Note: * p<0.05, ** p<0.01, *** p<0.001

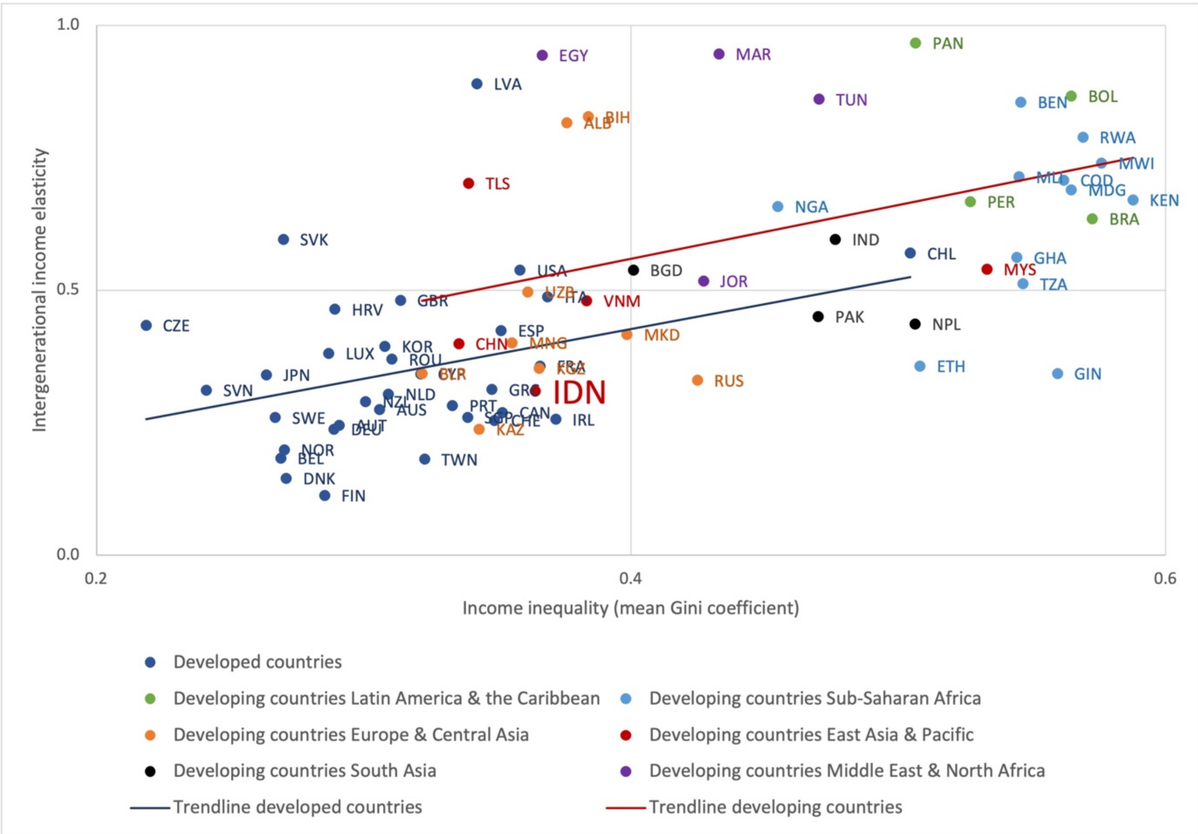
Source: authors' estimates based on IFLS.

4.3 International comparison of intergenerational mobility

The indices reported above permit comparison (with caveats) of mobility rates across countries. Two approaches for international comparison of mobility can be pursued. The first compares the results of independently conducted mobility studies from different countries, such as in Blanden (2009), Corak (2006), and Narayan et al. (2018). The second method analyses a sample containing data from all the different countries to be compared (comparators) and hence applies the same treatment to the data of each country. Studies applying the second approach include those of Björklund and Jäntti (1997), Couch and Dunn (1997), Grawe (2004), and Jäntti et al. (2006). No matter which method is applied, international comparisons of intergenerational income mobility are complicated for at least two reasons. First, most mobility measures are highly sensitive to the specific data definitions and data collection procedures applied. Patterns emerging from cross-country comparison may therefore largely reflect variations in data structure, measurement, and statistical approaches rather than genuine differences in intergenerational mobility. Second, as Jäntti et al. (2006) argue, there is no single objective summary measure of intergenerational mobility. In short, to conduct a meaningful comparison, it is not sufficient to compare the final index value of estimates. Certain information is needed to ensure that estimates of various countries are indeed comparable. The estimates should result from the same kind of copula; age restrictions should be applied in a similar way if not congruent; and equivalent definitions of economic outcomes should be used.

Figure 1 shows the Gini versus IGE of income.[5] We add our Indonesia estimate to depict Indonesia's position in the Great Gatsby Curve and enable an international comparison of β.[6] We depict the Great Gatsby Curve based on observations of IGE of income for father–son copulas (for consistent comparisons) in almost 60 developing countries and 18 developed countries from the World Bank's Global Database on Intergenerational Mobility (GDIM) (2018 version).[7]

Figure 1: IGE of income elasticity (father–son) versus mean Gini, year 0–20: developed countries (blue line of best fit) versus developing countries (red line of best fit)



Note: classification into developing/developed country according to World Bank current income group classification (HIC/non-HIC). Three outliers are not plotted on the figure (Colombia, Sri Lanka, South Africa).

Source: authors' estimates based on World Bank GDIM (2018) and UNU-WIDER (2021) WIID Companion. Indonesia estimate based authors' calculation using IFLS (father–son copulas from the birth cohort 1970).

We plot the IGE estimates for father–son against the Gini from the UNU-WIDER (2021) WIID Companion data set. To plot inequality, we calculate the means of all Gini observations within the period ranging from year 0 (birth of son) to year 20 (when it is assumed the son is economically

[5] Narayan et al. (2018: 114) also present plots using the GDIM database and separating developing economies and high-income countries using the 1960s cohort, IGE of income and '2000s data' for the Gini drawn from World Development Indicators (rather than UNU-WIDER WIID Companion). They note in brief that the correlation (of IGE of income with inequality) is stronger (the curve is steeper) in developing countries than in HICs.

[6] The estimates for other countries are comparable to those computed in our paper according to the three criteria outlined here (same kind of copula, age, and economic definitions). See also Narayan et al. (2018: 23, 141, figure 4.2) and OECD (2008: 213).

[7] The 2021 version of GDIM focuses on education and mobility only.

independent).[8] The dates of the periods were selected to include at least 20 years in the range if data availability allowed it.[9] Different colours are used to distinguish developed countries (high-income countries by current World Bank classification) from developing countries (low- and middle-income countries), the latter also being colour coded by region. As is evident, such comparisons are—at best—crude. Further, as Iversen et al. (2021) note, it makes more sense to plot the IGE of income against a measure of the inequality of opportunity. In Figure 1 we have plotted IGE of income against inequality of outcome to follow earlier studies on the Great Gatsby curve and Kanbur (2021). Kanbur notes that there are positive and normative reasons to retain a focus on inequality of outcome: first, because inequality of outcome may itself be a determinant of intergenerational mobility; and second, because objectives such as the equality of educational outcomes are related to mobility. Further support for these arguments is presented in Narayan et al. (2018).

A set of stylized facts are worth noting. First, the data show that IGE of income is positively associated—in general—with income inequality. Second, the data show that developed countries are largely situated in the bottom left quadrant, with lower inequality and a lower IGE of income. There are also several transition economies. Third, in contrast to developed countries, developing countries are situated in, or in the vicinity of, the top right quadrant, which represents higher income inequality and higher IGE of income. In other words, with a few notable exceptions, the Great Gatsby Curve shows that developed countries exhibit higher mobility and less inequality, and that the opposite is true, in general, for developing countries.

Among developed countries, the US, Slovakia, and Latvia are outliers, exhibiting deviation in one or both indicators. Indonesia, when our estimate is added, and Kazakhstan are two exceptions for developing countries, as they have lower IGE. Furthermore, if we draw a linear line of best fit for each country grouping, we see that at any given level of income inequality the IGE of income is likely to be higher (and thus mobility worse) in a developing country than in a developed country. We discuss these issues further in Sakri et al. (forthcoming).

In sum, there is an association between IGE and income inequality. However, countries tend to be segregated into two distinct groups: developed countries (HICs) with lower income inequality and lower IGE of income and. conversely, developing countries with—in general—higher income inequality and higher IGE of income.

5 Gender, geography, and generations and mobility

5.1 The nonlinearity of mobility

How much do estimates of IGE of income differ by gender, geography, or generation (birth period) within a country? IGE estimates (β) summarize the relationship between two generations at the mean value of their outcomes. However, according to Black and Devereux (2011), there is no reason to assume that β is linear for different parts of the children's income distribution. Therefore, it is necessary to check this nonlinearity hypothesis before continuing with further dynamics analysis. A recent theoretical contribution by Becker et al. (2018) on nonlinearity of parents–children relationships predicts that intergenerational mobility will not be constant across

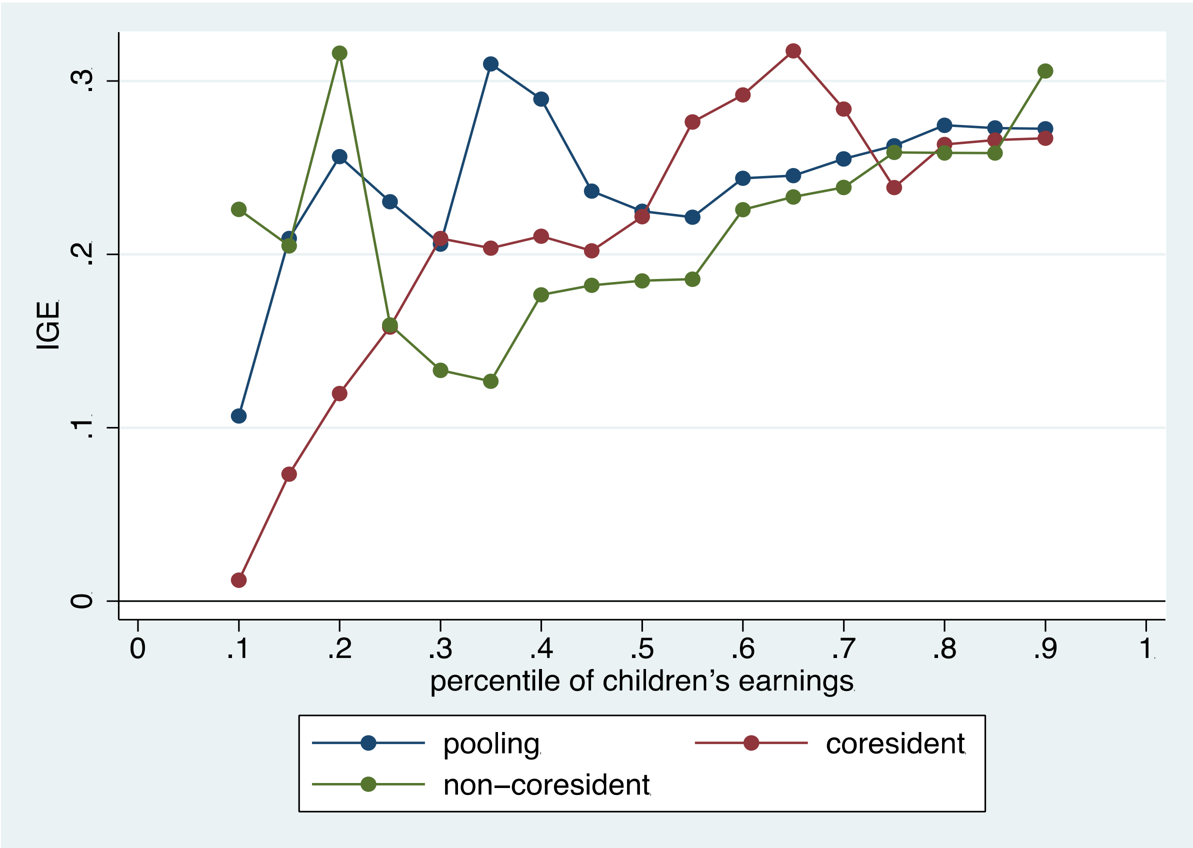
[8] Thus, in contrast to Narayan et al. (2018), we use the Gini and GNI shares for the years when the children were still nurtured by their parents. Other authors such as Corak (2013) employ the Gini for the year 1980.

[9] If no data were available for the year of childbirth, the earliest available year closest to 1990 was chosen.

the entire income distribution. This is partly because low-income parents are likely to experience credit constraints while wealthier parents have a greater chance to invest in their children's human capital due to higher returns on those investments. This will lead to higher intergenerational persistence at the top and the bottom of the parental income distribution.

To check the nonlinearity hypothesis in Indonesia, IGE estimates from pooling, coresident, and non-coresident samples were tested using the RIF-regression technique.[10] Figure 2 depicts IGE across different quantiles of the children's earnings distribution for the three different subsets of the sample. It provides evidence that high persistence of earnings is indeed found at the top of the distribution.

Figure 2: Nonlinearity of IGE estimates



Source: authors' estimate based on IFLS.

In all samples, children in the top 30 per cent by income consistently get a higher IGE score. This means that their outcomes are strongly correlated with those of their parents, who are at the top of the income distribution as well. Furthermore, the non-coresident sample almost perfectly follows the hypothesis that there is relatively higher IGE at the bottom end of the distribution compared with the middle; hence it is mimicking a J-shaped curve. However, the pooling sample, by tendency, demonstrates higher IGE at both the lower tail and the centre. Hence, it does not follow the nonlinearity hypothesis as smoothly as the non-coresident sample. Likewise, the coresident sample does not follow the hypothesis at the bottom of the distribution. In short, although the curves do not perfectly match the hypothesis of high persistency (β) at the bottom and top tails of the income distribution, Figure 2 does provide evidence of the nonlinearity of β.

[10] See Firpo et al. (2009).

5.2 Dynamics of intergenerational mobility in Indonesia

Existing mobility studies have shown that the mobility rates of the same generation in the same country differ by gender (male vs female), temporal aspects (birth cohort), and regional characteristics (specific geographic/sub-national areas). Motivated by these findings, we analyse these three factors that potentially alter the mobility rates of particular individuals in Indonesia. In terms of gender and intergenerational mobility (or in other development processes), there are factors (e.g. social norms, discrimination practices) that affect the outcomes of women. Cohort difference is considered to determine temporal dynamics of mobility. Lastly, the interest in potential mobility differences due to parental regional origin is motivated by the fact that Indonesia is a large archipelagic country. Hence, there is no reason to assume that the mobility rate is the same across diverse regions.

Furthermore, given that the assessment of nonlinearity of mobility in Section 4.4 was based on the pooling sample, which does not distinguish by gender, birth cohort, or regional origin, it is logical to assume that this nonlinearity to some extent stems from these differences. The following sections will analyse differences due to each of these dimensions.

5.3 Gender and mobility

Table 6 reports summary statistics of the sample split by gender (5,452 for males and 3,993 for females). The age gap between parents and their children appears unaffected by the gender of the latter. Additionally, the mean ages of both samples are around 43 years for parents and 34 years for children. Age thus represents only a minimal risk of biased mobility estimates drawn from the split gender sample compared with those stemming from the pooling sample. The number of data points used for averaging earnings values of respondents is similar to that of the previous analysis, namely around three for parents and two for children.

Table 6: Sample for gender difference analysis

Variable	N	Min.	Max.	P50	Mean	SD
Male						
EarningG0	5,163	7.94	16.38	12.62	12.58	1.10
EarningG1	5,452	7.97	16.41	13.43	13.36	1.06
DataG0	5,452	0.00	5.00	3.00	2.81	1.26
DataG1	5,452	1.00	5.00	2.00	1.99	1.01
AgeG0	5,452	18.50	88.00	42.00	42.89	10.53
AgeG1	5,452	21.00	79.00	33.00	34.09	7.87
Female						
EarningG0	3,726	7.29	15.66	12.64	12.60	1.10
EarningG1	3,993	8.28	16.69	12.86	12.82	1.20
DataG0	3,993	0.00	5.00	3.00	2.78	1.30
DataG1	3,993	1.00	5.00	1.00	1.66	0.87
AgeG0	3,993	19.00	86.00	41.50	43.12	10.71
AgeG1	3,993	21.00	80.00	33.00	34.46	8.45

Source: authors' estimates based on IFLS.

Table 7 reports differences in IGE between male and female children, both estimated by β and ρ. β estimates suggest that males have higher mobility rates than their female counterparts. The difference amounts to around five percentage points and is significant, as lower- and upper-bound estimates are almost perfectly separated. However, ρ estimates indicate that male and female children experience equal mobility levels.

Table 7: Comparison of IGE estimates: male vs female

IGE estimates	β		ρ	
Gender	**Male**	**Female**	**Male**	**Female**
EarningG0	0.289***	0.348***	0.313***	0.310***
SE	(0.013)	(0.018)	(0.013)	(0.015)
Lower-bound	0.264	0.314	0.288	0.281
Upper-bound	0.315	0.383	0.338	0.339
Constant	9.731***	8.448***	34.718***	34.881***
SE	(0.163)	(0.224)	(0.743)	(0.865)
R-squared	0.090	0.102	0.099	0.098
N	5,163	3,726	5,452	3,993

Note: * p<0.05, ** p<0.01, *** p<0.001

Source: authors' estimates based on IFLS.

5.4 Temporal dynamics of mobility

Table 8 reports summary statistics of the pre-millennial and millennial sample. The number of observations is split into 4,153 for the former (born before 1980) and 5,292 for the latter (born in or after 1980). The number of observation points on earnings data is between two and three for both earnings of children and their parents. Hence, it can be assumed that the observations are relatively robust to transitory income shocks.

Table 8: Sample for temporal dynamics analysis

Variable	N	Min.	Max.	P50	Mean	SD
Pre-millennials						
EarningG0	3,644	7.29	16.16	12.32	12.25	1.17
EarningG1	4,153	8.21	16.69	13.03	12.99	1.12
DataG0	4,153	0.00	5.00	2.50	2.30	1.35
DataG1	4,153	1.00	5.00	2.00	2.32	1.07
AgeG0	4,153	26.00	88.00	50.00	50.08	9.37
AgeG1	4,153	35.00	80.00	40.00	41.52	6.07
Millennials						
EarningG0	5,245	8.26	16.22	12.84	12.83	0.99
EarningG1	5,292	7.97	16.74	13.37	13.26	1.17
DataG0	5,292	0.00	5.00	3.00	3.18	1.06
DataG1	5,292	1.00	4.00	1.00	1.49	0.68
AgeG0	5,292	18.50	72.50	36.50	37.42	7.84
AgeG1	5,292	21.00	34.00	29.00	28.54	3.84

Source: authors' estimates based on IFLS.

The mean age of parents of pre-millennials was 50 years and their children were on average 42 years old when their incomes were recorded. In contrast, the mean age of parents of millennial children was 37 years and their children were on average 29 years old. These ages are still sufficiently close to the midpoint of their respective careers. Hence, the likelihood of lifecycle bias in measured earnings is minimal, though it should be noted (see below).

Table 9 reports a significant difference between pre-millennials' and millennials' mobility rates. According to β estimates (second and third row), at a 95 per cent confidence interval, the lower and upper bounds of the two groups are clearly far apart. This gives us confidence to infer that the mobility rate is declining by a factor of around seven percentage points when comparing pre-millennials with millennials. This conclusion is supported by rank-rank specification estimates, ρ

(fourth and fifth row). It was shown in the previous section that ρ yields lower estimates of IGE than β. Hence, it should not be surprising that ρ estimates are lower than β estimates for both ends. However, the most important insight from ρ in this case is that it corroborates a difference in mobility rates between pre-millennials and millennials.

Table 9: Comparison of IGE estimates: pre-millennials vs millennials

IGE estimates	β		ρ	
Generation	**Pre-millennials**	**Millennials**	**Pre-millennials**	**Millennials**
EarningG0	0.264***	0.337***	0.249***	0.324***
SE	(0.016)	(0.017)	(0.015)	(0.015)
Lower-bound	0.232	0.305	0.220	0.295
Upper-bound	0.295	0.370	0.277	0.354
Constant	9.773***	8.928***	36.278***	34.857***
SE	(0.195)	(0.213)	(0.703)	(0.956)
R-squared	0.075	0.082	0.068	0.084
N	3,644	5,245	4,153	5,292

Note: * $p<0.05$, ** $p<0.01$, *** $p<0.001$

Source: authors' estimate based on IFLS.

The first probable cause of this decline in mobility is the age difference between the two groups at the moment of measurement. Although it has been mentioned that the mobility estimates are assumed to have minimal lifecycle bias since the mean age of the children was high enough for them to be employed when their earnings were measured, the effect might still become apparent. While pre-millennials (41.5 years old on average) might be approaching the income peak of their working life, millennials (28.5 years of age) are in earlier stages of their working life. Likewise, the parents of millennials (mean age 37.4) might be at their peak whereas the productivity of the parents of pre-millennials (50.1) might be post-peak.

To assess the effect of age on mobility estimates, this research follows other studies such as Deng et al. (2013) and Ferreira and Veloso (2006). Consequently, we include age and squared age of both children and parents in the two equations (1) and (2) mentioned earlier in this paper. Table 10 reports the results and illustrates that the age of neither parents nor children affects mobility estimates, irrespective of whether β or ρ is measured. Furthermore, the difference between pre-millennials and millennials remains approximately the same, which means that the former experience lower IGE (higher mobility) than the latter. This implies that the findings of declining mobility rates between pre-millennials and millennials most likely do not stem from data construction.

Table 10: Age-adjusted comparison of IGE estimates: pre-millennials vs millennials

IGE estimates	β		ρ	
Generation	**Pre-millennials**	**Millennials**	**Pre-millennials**	**Millennials**
EarningG0	0.262***	0.340***	0.250***	0.325***
SE	(0.018)	(0.018)	(0.017)	(0.016)
Lower-bound	0.227	0.306	0.217	0.294
Upper-bound	0.297	0.374	0.283	0.356
AgeG0	0.021	0.067***	0.766*	1.687***
SE	(0.021)	(0.015)	(0.391)	(0.369)
SqAgeG0	-0.000	-0.001***	-0.007	-0.018***
SE	(0.000)	(0.000)	(0.004)	(0.005)
AgeG1	0.049	0.094	1.971***	1.680
SE	(0.038)	(0.063)	(0.551)	(1.566)
SqAgeG1	-0.001	-0.002*	-0.022***	-0.046
SE	(0.000)	(0.001)	(0.006)	(0.028)
Constant	8.236***	6.609***	-26.640*	-11.763
SE	(0.858)	(0.907)	(12.917)	(22.008)
R-squared	0.075	0.091	0.073	0.095
N	3,644	5,245	4,153	5,292

Note: * p<0.05, ** p<0.01, *** p<0.001

Source: authors' estimate based on IFLS.

5.5 Regional differences in mobility

Table 11 reports summary statistics of the sample split by the 13 provinces of parental origin.

Table 11: Sample for provincial dynamics analysis

Province	N	EarningG0	EarningG1	DataG0	DataG1	AgeG0	AgeG1
Kalsel	392	12.63	13.21	2.85	1.87	41.66	33.26
Lampung	423	12.52	13.02	2.92	1.76	41.74	32.18
Bali	497	12.79	13.25	3.28	1.93	40.73	33.16
Sulsel	501	12.38	12.66	2.64	1.70	43.05	34.52
Sumsel	502	12.81	13.36	2.58	1.69	42.47	33.27
Sumbar	503	12.66	13.38	2.80	1.74	43.03	34.17
Yogyakarta	520	12.24	13.11	3.04	2.18	47.16	37.18
NTB	582	12.49	12.81	2.97	1.77	43.41	33.39
Sumut	785	12.76	13.26	2.74	1.62	42.28	33.57
Jakarta	943	13.21	13.79	2.39	1.83	44.14	35.94
Jatim	1,152	12.37	12.84	2.83	1.90	42.47	34.67
Jateng	1,225	12.30	12.95	3.05	1.96	43.36	34.06
Jabar	1,420	12.63	13.18	2.59	1.93	42.68	34.14
Indonesia	9,445	12.59	13.14	2.80	1.85	42.99	34.25

Note: NTB = Nusa Tenggara Barat.

Source: authors' estimates based on IFLS.

The number of observations per province varies. For some provinces the number is very low, such as for Kalimatan Selatan (392 observations). For others the number is reasonably high, for instance for Jawa Barat (1,420 observations). However, out of 13 provinces, only three have more than a thousand observations. This raises concerns about the quality of mobility estimates resulting from such a low number of observations; hence any inferences should be treated with caution. As noted before, many mobility studies suffer from questions over the number of observations in their sample.

Nevertheless, Table 11 indicates that children's mean ages are similar across all provinces, namely between 32 years (in Lampung) and 37 years (in DI Yogyakarta). The mean age of parents ranges from 41 in Bali to 47 years in DI Yogyakarta. Thus, following the distribution of mean age in the previous analysis, it is reasonable to assume that mobility estimates at province level should not be affected by lifecycle bias. Consistent with previous sub-samples, like male vs female, pre-millennials vs millennials, and coresident vs non-coresident, earnings data are averaged from typically three observations for parents and two for their children. This is a distribution of the sample that enables mobility analysis to continue at provincial level.

Table 12 compares relative mobility measured by β (upper half of the table) and by ρ (bottom half). For each measure, provinces were sorted from smallest to highest preferred estimates (second column).

Table 12: Provincial mobility estimates

Province	EarningG0	SE	Lower-bound	Upper-bound	Constant	SE	R-squared	N
β estimates								
Sulsel	0.18***	-0.05	0.083	0.278	10.46***	-0.62	0.03	462
Yogyakarta	0.203***	-0.04	0.129	0.277	10.64***	-0.46	0.06	502
Lampung	0.204***	-0.06	0.096	0.312	10.45***	-0.68	0.03	4411
Jakarta	0.224***	-0.03	0.162	0.285	10.87***	-0.42	0.05	860
Jatim	0.229***	-0.03	0.171	0.287	10.02***	-0.37	0.05	1071
Kalsel	0.266***	-0.05	0.167	0.364	9.84***	-0.64	0.07	379
Bali	0.285***	-0.05	0.187	0.382	9.62***	-0.64	0.07	489
Jateng	0.294***	-0.03	0.236	0.352	9.34***	-0.37	0.08	1176
Sumsel	0.298***	-0.05	0.204	0.391	9.58***	-0.62	0.08	479
Sumut	0.309***	-0.04	0.231	0.387	9.34***	-0.51	0.07	744
Indonesia	0.31***	-0.01	0.289	0.332	9.25***	-0.14	0.09	8889
Jabar	0.321***	-0.03	0.263	0.38	9.11***	-0.38	0.09	1286
Sumbar	0.324***	-0.05	0.221	0.426	9.32***	-0.67	0.08	475
NTB	0.328***	-0.04	0.24	0.415	8.75***	-0.56	0.09	555
ρ estimates								
Lampung	0.208***	-0.05	0.108	0.308	36.22***	-2.70	0.04	423
Jakarta	0.212***	-0.03	0.158	0.266	53.72***	-2.01	0.06	943
Sulsel	0.217***	-0.05	0.129	0.306	30.39***	-2.30	0.05	501
Jabar	0.231***	-0.02	0.183	0.278	40.17***	-1.37	0.06	1420
Jatim	0.234***	-0.03	0.18	0.288	32.13***	-1.38	0.06	1152
Kalsel	0.235***	-0.05	0.14	0.329	39.96***	-2.92	0.06	392
Yogyakarta	0.258***	-0.04	0.178	0.338	37.92***	-2.08	0.07	520
Sumut	0.276***	-0.04	0.206	0.346	39.04***	-2.19	0.07	785
Jateng	0.286***	-0.03	0.228	0.343	32.43***	-1.48	0.07	1225
Indonesia	0.299***	-0.01	0.28	0.319	35.42***	-0.56	0.09	9445
Sumbar	0.321***	-0.04	0.235	0.407	39.97***	-2.66	0.09	503
NTB	0.321***	-0.04	0.243	0.4	26.61***	-2.11	0.1	582
Bali	0.323***	-0.05	0.234	0.412	35.04***	-2.85	0.09	497
Sumsel	0.342***	-0.04	0.262	0.422	37.12***	-2.61	0.12	502

Note: * p<0.05, ** p<0.01, *** p<0.001

Source: authors' estimates based on IFLS.

his paper has reported two measures of relative mobility for Indonesia, namely IGE based on log-log regression of children's outcomes on their parents' outcomes (β) and rank-rank specification (ρ). Both were estimated from 9,445 matching pairs of children and their parents (pooling sample). The estimates proved robust to possible transitory shocks and lifecycle bias that might skew measured incomes, and to coresidency bias due to the use of household surveys.

To provide a meaningful interpretation of mobility rates, Indonesia's mobility estimates were compared with readily available international estimates. To ensure comparability, the estimation of Indonesia's mobility rate followed and mimicked the approaches underlying the existing estimates in terms of the type of copulas analysed (fathers–sons instead of parents–children), age of sons and fathers when income is measured, and economic outcomes used (i.e. earnings instead of family income). Several key findings emerged from this comparison.

First, Indonesia was found to be an exception to the Great Gatsby Curve, which depicts the relationship between IGE (father–son copulas) and income inequality. Generally, the Great Gatsby Curve illustrates that countries with high inequality exhibit low mobility (high IGE scores) and, conversely, countries with low inequality demonstrate high mobility. The former case mostly relates to developing countries, whereas the latter is mostly observed in developed countries. Indonesia, which is a developing country, emerges as one of the few exceptions in terms of the relationship between inequality and mobility depicted in the Great Gatsby Curve since it exhibits a low IGE (high mobility) compared with other developing countries. One factor potentially explaining this comparatively high mobility in Indonesia is the progressivity of government investments in education during the 1970s until the early 1990s, which was a period of publicly funded basic education. This period corresponds to the childhood era of the analysed children's cohort of 1970. This implies that these children benefited from the programme, which increased their education levels; as a result, their incomes significantly deviate from their respective parents' incomes.

Subsequently, we examined the nonlinearity hypothesis of IGE, which predicts that intergenerational mobility will not be constant across the entire income distribution. This is partly because low-income parents are likely to experience credit constraints while wealthier parents have the possibility to invest more in their children's human capital. This is assumed to lead to higher intergenerational persistence at the top and the bottom of the parental income distribution. To check the nonlinearity hypothesis in Indonesia, IGE estimates from pooling, coresident, and non-coresident samples were tested using the RIF-regression technique. It was found that in all sub-samples, children with an income position in the top 30 per cent consistently obtained higher IGE scores. This means that their outcomes are strongly correlated with the outcomes of their parents, who are at the top of the income distribution as well. The non-coresident sample almost perfectly follows the nonlinearity hypothesis and reveals a relatively higher IGE at the bottom compared with the middle, hence mimicking a J-shaped curve. However, the pooling sample demonstrates a tendency of a higher IGE at the low tail as well as in the centre; hence it does not follow the hypothesis as smoothly as the non-coresident sample. Likewise, the coresident sample does not follow the nonlinearity hypothesis at its bottom end. It is logical to assume that this nonlinearity to some extent stems from various differentiating factors, including gender, birth cohort, and parental regional origin.

Having investigated its nonlinearity, we analysed differences in mobility due to the gender of the children, their generation (birth cohort), and parental provincial origin. Splitting the sample by gender revealed that male children tend to be more mobile than female children by a factor of almost six percentage points (0.289 vs 0.348). To check on mobility differences due to birth cohort, the pooling sample was split into two groups, namely pre-millennials (born before 1980) and millennials (born in or after 1980). It became evident that pre-millennials exhibit higher mobility (lower IGE) than their counterparts by a factor of 7.3 percentage points (0.264 vs 0.337). This difference proved robust to age differences between the two groups; hence it is probably caused by factors external to data construction. Lastly, to assess regional differences of mobility, the pooling sample was split by the 13 provinces where the parents originated. The sub-sample contains an unbalanced number of observations ranging from as low as 392 in Kalimatan Selatan to as high as 1,420 in Jawa Barat. The small number of observations in some cases likely presents a caveat to mobility analysis at the provincial level. Yet there is no big difference between the provincial and the pooling samples in terms of the number of data points used to average earnings as well as the mean ages of children and parents. Relative mobility (β) across provinces is spread by more than 14 percentage points, from the smallest in Sulawesi Selatan (0.18) to the highest in NTB (0.328). In sum, the presence of nonlinearity of mobility and differences due to gender, birth cohort, and provincial origin is strongly supported by the data.

References

Atkinson, A.B. (1980). 'On intergenerational income mobility in Britain'. *Journal of Post Keynesian Economics*, 3(2): 194–218. https://doi.org/10.1080/01603477.1980.11489214

Atkinson, A.B., C.G. Trinder, and A.K. Maynard (1978). 'Evidence on intergenerational income mobility in Britain'. *Economics Letters*, 1(4), 383–88. https://doi.org/10.1016/0165-1765(78)90017-4

Becker, G.S., and N. Tomes (1986). 'Human capital and the rise and fall of families'. *Journal of Labor Economics*, 4(3, Part 2): 1–39. https://doi.org/10.1086/298118

Becker, G.S., S.D. Kominers, K.M. Murphy, and J.L. Spenkuch (2018). 'A theory of intergenerational mobility'. *Journal of Political Economy*, 126(S1): 7–25. https://doi.org/10.1086/698759

Björklund, A., and M. Jäntti (1997). 'Intergenerational income mobility in Sweden compared to the United States'. *The American Economic Review*, 87(5): 1009–18.

Björklund, A., J. Roine, and D. Waldenström (2012). 'Intergenerational top income mobility in Sweden: capitalist dynasties in the land of equal opportunity?'. *Journal of Public Economics*, 96(5–6): 474–84. https://doi.org/10.1016/j.jpubeco.2012.02.003

Black, S.E., and P.J. Devereux (2011). 'Recent developments in intergenerational mobility'. In O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, pp. 1487–541. Amsterdam: Elsevier. https://doi.org/10.1016/S0169-7218(11)02414-2

Blanden, J. (2009). 'How much can we learn from international comparisons of intergenerational mobility?'. CEE Discussion Paper 111. London: Centre for the Economics of Education, London School of Economics and Political Science.

Bratberg, E., J. Davis, B. Mazumder, M. Nybom, D.D. Schnitzlein, and K. Vaage (2017). 'A comparison of intergenerational mobility curves in Germany, Norway, Sweden, and the US'. *The Scandinavian Journal of Economics*, 119(1): 72–101. https://doi.org/10.1111/sjoe.12197

Bratsberg, B., K. Røed, O. Raaum, R. Naylor, M. Jäntti, T. Eriksson, and E. Österbacka (2007). 'Nonlinearities in intergenerational earnings mobility: consequences for cross-country comparisons'. *The Economic Journal*, 117(519): 72–92. https://doi.org/10.1111/j.1468-0297.2007.02036.x

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014a). 'Where is the land of opportunity? The geography of intergenerational mobility in the United States'. *The Quarterly Journal of Economics*, 129(4): 1553–623. https://doi.org/10.1093/qje/qju022

Chetty, R., N. Hendren, P. Kline, E. Saez, and N. Turner (2014b). 'Is the United States still the land of opportunity? Recent trends in intergenerational mobility'. *American Economic Review*, 104(5): 141–47. https://doi.org/10.1257/aer.104.5.141

Corak, M. (2006). 'Do poor children become poor adults? Lessons from a cross-country comparison of generational earnings mobility'. *Research on Economic Inequality*, 13(1): 143–88. https://doi.org/10.1016/S1049-2585(06)13006-9

Corak, M. (2013). 'Income inequality, equality of opportunity, and intergenerational mobility'. *Journal of Economic Perspectives*, 27(3): 79–102. https://doi.org/10.1257/jep.27.3.79

Corak, M., M.J. Lindquist, and B. Mazumder (2014). 'A comparison of upward and downward intergenerational mobility in Canada, Sweden and the United States'. *Labour Economics*, 30: 185–200. https://doi.org/10.1016/j.labeco.2014.03.013

Couch, K.A., and T.A. Dunn (1997). 'Intergenerational correlations in labor market status: a comparison of the United States and Germany'. *Journal of Human Resources*, 32(1): 210–32. https://doi.org/10.2307/146246

Dartanto, T., F.R. Moeis, and S. Otsubo (2019). 'Intragenerational economic mobility in Indonesia: a transition from poverty to the middle class in 1993–2014'. *Bulletin of Indonesian Economic Studies*, 56(2): 193–224. https://doi.org/10.1080/00074918.2019.1657795

Deng, Q., B. Gustafsson, and S. Li (2013). 'Intergenerational income persistence in urban China'. *Review of Income and Wealth*, 59(3): 416–36. https://doi.org/10.1111/roiw.12034

Deutscher, N., and B. Mazumder (forthcoming). 'Measuring intergenerational income mobility: a synthesis of approaches'. *Journal of Economic Literature*.

DiPrete, T.A. (2020). 'The impact of inequality on intergenerational mobility'. *Annual Review of Sociology*, 46: 379–98. https://doi.org/10.1146/annurev-soc-121919-054814

Durlauf, S.N., A. Kourtellos, and C. Ming Tan (2021). 'The Great Gatsby Curve'. Working Paper 43. New York: Stone Centre on Socio-Economic Inequality. https://doi.org/10.31235/osf.io/mrw9y

Emran, M.S., W. Greene, and F. Shilpi (2016). 'When measure matters: coresidency, truncation bias, and intergenerational mobility in developing countries'. Policy Research Working Paper 7608. Washington, DC: The World Bank. https://doi.org/10.1596/1813-9450-7608

Ferreira, S.G., and F.A. Veloso (2006). 'Intergenerational mobility of wages in Brazil'. *Brazilian Review of Econometrics*, 26(2): 181–211. https://doi.org/10.12660/bre.v26n22006.1576

Fields, G.S. (2005). 'The many facets of economic mobility'. In M. McGillivray (ed.), *Inequality, poverty and well-being*, pp. 123–42. Basingstoke: Palgrave Macmillan. https://doi.org/10.1057/9780230625594_6

Fields, GS. (2021). 'Exploring concepts of social mobility'. In V. Iversen, A. Krishna, and K. Sen (eds), *Social mobility in developing countries: concepts, methods, and determinants*. Oxford: Oxford University Press; Helsinki: United Nations University World Institute for Development Economics Research (UNU-WIDER).

Fields, G.S., and E.A. Ok (1999). 'The measurement of income mobility: an introduction to the literature'. In J. Silber (ed.), *Handbook of income inequality measurement*, pp. 557–98. Boston: Kluwer Academic Publishers. https://doi.org/10.1007/978-94-011-4413-1_20

Firpo, S., P. Nicole, M. Fortin, and T, Lemieux (2009). 'Unconditional quantile regressions'. *Econometrica*, 77(3): 953–73. https://doi.org/10.3982/ECTA6822

Friedman, J., and J. Levinsohn (2001). 'The distributional impacts of Indonesia's financial crisis on household welfare: a "rapid response" methodology'. NBER Working Paper 8564. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.1093/wber/lhf001

Glass, D.V. (1954). *Social Mobility in Britain*. London: Routledge.

Grawe, N.D. (2004). 'Intergenerational mobility for whom? The experience of high- and low-earning sons in international perspective'. In M. Corak (ed.), *Generational income mobility in North America and Europe*, pp. 58–89. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511492549.005

Grawe, N.D., and C.B. Mulligan (2002). 'Economic interpretations of intergenerational correlations'. *Journal of Economic Perspectives*, 16(3): 45–58. https://doi.org/10.1257/089533002760278703

Gregg, P., L. Macmillan, and C. Vittori (2019). 'Intergenerational income mobility: access to top jobs, the low-pay no-pay cycle and the role of education in a common framework'. *Journal of Population Economics*, 32(2): 501–28. https://doi.org/10.1007/s00148-018-0722-z

ILO (2007). *International Standard Classification of Occupations (ISCO-08): structure, group definitions and correspondence tables*. Geneva: International Labour Organization.

Iversen, V., A. Krishna, and K. Sen (2021). 'The state of knowledge about social mobility in the developing world'. In V. Iversen, A. Krishna, and K. Sen (2021). *The state of knowledge about social mobility in the developing world: concepts, methods, determinants*. Oxford: OUP. https://doi.org/10.1093/oso/9780192896858.001.0001

Jäntti, M., and S.P. Jenkins (2015). 'Income mobility'. In A.B. Atkinson and F. Bourguignon (eds), *Handbook of Income Distribution* (pp. 807–936). Amsterdam: Elsevier. https://doi.org/10.1016/B978-0-444-59428-0.00011-4

Jäntti, M., et al. (2006). 'American exceptionalism in a new light: a comparison of intergenerational earnings mobility in the Nordic Countries, the United Kingdom and the United States'. IZA Discussion Paper 1938. Bonn: Institute of Labor Economics (IZA). https://doi.org/10.2139/ssrn.878675

Kanbur, R. (2021). 'In praise of snapshots'. In V. Iversen, A. Krishna, and K. Sen (eds), *Social mobility in developing countries: concepts, methods, and determinants*. Oxford: Oxford University Press; Helsinki: United Nations University World Institute for Development Economics Research (UNU-WIDER).

Krueger, A.B. (2012). 'The rise and consequences of inequality in the United States' [Transcript]. Washington, DC: Center for American Progress. Available at: http://bit.do/americanprogress-org-2012-01-krueger (accessed 21 April 2020).

Narayan, A., et al. (2018). 'Fair progress? Economic mobility across generations around the world'. Washington, DC: The World Bank. https://doi.org/10.1596/978-1-4648-1210-1

Nybom, M., and J. Stuhler (2015). 'Biases in standard measures of intergenerational income dependence'. *Journal of Human Resources*, 52(3): 800–25. https://doi.org/10.3368/jhr.52.3.0715-7290R

Nybom, M., and J. Stuhler (2016). 'Heterogeneous income profiles and lifecycle bias in intergenerational mobility estimation'. *Journal of Human Resources*, 51(1): 239–68. https://doi.org/10.3368/jhr.51.1.239

OECD (2008). *Growing unequal: income distribution and poverty in OECD countries*. Paris: Organisation for Economic Co-Operation and Development. https://doi.org/10.1787/9789264044197-en

OECD (2011). *Divided we stand: why inequality keeps rising*. Paris: Organisation for Economic Co-Operation and Development. https://doi.org/10.1787/9789264119536-en

Pekkarinen, T., K.G. Salvanes, and M. Sarvimäki (2017). 'The evolution of social mobility: Norway during the twentieth century'. *The Scandinavian Journal of Economics*, 119(1): 5–33. https://doi.org/10.1111/sjoe.12205

Piraino, P. (2021). 'Drivers of mobility in the Global South'. In V. Iversen, A. Krishna, and K. Sen (eds), *Social mobility in developing countries: concepts, methods, and determinants*. Oxford: Oxford University Press; Helsinki: United Nations University World Institute for Development Economics Research (UNU-WIDER)

Pradhan, M., A. Suryahadi, S. Sumarto, and L.H. Pritchett (2000). 'Measurements of poverty in Indonesia: 1996, 1999, and beyond'. Policy Research Working Paper 2438. Washington, DC: The World Bank. https://doi.org/10.1596/1813-9450-2438

Quiñones, E.J., A.P.D. O-Campos, and C. Rodríguez-Alas (2009). *Methodology for creating the RIGA-L database*. Rome: Food and Agriculture Organization.

Roemer, J.E. (2004). 'Equal opportunity and inter-generational mobility: going beyond intergenerational income transition matrices'. In M. Corak (ed.), *Generational income mobility in North America and Europe*. Cambridge: Cambridge University Press.

Sakri, D. (2019). 'Intergenerational income mobility in Indonesia, 1993–2014'. PhD Thesis. King's College London.

Sakri, D., A. Sumner, and A. Yusuf (forthcoming). 'Inequality and intergenerational dynamics in developing countries'. Mimeograph.

Savegnago, M. (2016). 'Igmobil: a command for intergenerational mobility analysis in Stata'. *The Stata Journal*, 16(2): 386–402. https://doi.org/10.1177/1536867X1601600207

Solon, G. (1992). 'Intergenerational income mobility in the United States'. *American Economic Review*, 82(3): 393–408.

Solon, G. (1999). 'Intergenerational mobility in the labor market'. In O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, pp. 1761–800. Amsterdam: Elsevier. https://doi.org/10.1016/S1573-4463(99)03010-2

Sorokin, P.A. (1927). *Social mobility*. New York: Harper & Row.

Strauss, J., K. Beegle, B. Sikoki, A. Dwiyanto, Y. Herawati, and F. Witoelar (2004). *The third wave of the Indonesia Family Life Survey (IFLS3): overview and field report*. Santa Monica: RAND Corporation.

Strauss, J., F. Witoelar, and B. Sikoki (2016). *User's guide for the Indonesia Family Life Survey, Wave 5: Volume 2*. Santa Monica: RAND Corporation. https://doi.org/10.7249/WR1143.2

UNU-WIDER (2021). 'World Income Inequality Database (WIID) Companion dataset'. Version 31 May 2021. https://doi.org/10.35188/UNU-WIDER/WIIDcomp-310521

Witoelar, F. (2009). *Note on the construction of the IFLS consumption expenditure aggregates*. Washington, DC: The World Bank.

Appendix: estimating earnings or personal income from the IFLS

Step 1: Defining the main variable and its covariates

Table A1 defines four variables (Variables 1–4) as covariates to the main variable (Variable 5). Although earnings is our main variable, Variables 1 to 4 must also be drawn from the IFLS in case they are needed to estimate or impute earnings data.

Table A1: Variables constructed from IFLS

No.	Variable name	Definition/remarks
1	Working time	Duration (total weeks per year) and frequency (total hours per week) of work
2	Occupation	Occupation and respective ISCO-08 classification
3	Employment	Type of employment
4	Sector	Nine-sector economy in which respondent works
5	Earnings	Individual income from three sources: net profit, salary, bonus

Source: authors' construction.

Variables 1 to 5 are related to individual employment activity. In the IFLS, questionnaires pertaining to those variables are recorded in Book 3 Section TK (employment). Section TK module B3A_TK1 screens the adult members of households with the following four questions:

TK01. What was your primary activity during the past week?

1. Working/trying to work/helping to earn income
2. Job searching
3. Attending school
4. Housekeeping
5. Retired
6. Sick/disabled
7. Other

TK02. Did you work/try to work/help to earn income for at least one hour of the week during the past week?

1. Yes
2. No

TK03. Do you have a job/business but were temporarily not working during the past week?

1. Yes
2. No

TK04. Did you work at a family-owned (farm or non-farm) business during the past week?

1. Yes
2. No

The data reported in this paper with regard to Variables 1 to 5 are extracted from respondents recorded in Book 3 Section TK who answered '1 (yes)' to at least one of Questions 02–04. With that screening step, the numbers of observations included from each wave of the IFLS were as follows: 9,762 (IFLS93), 11,964 (IFLS97), 17,317 (IFLS00), 20,166 (IFLS07), and 24,489 (IFLS14).

We will refer to these figures as the total population for Variables 1 to 5 and we will compare all non-reporting and reporting data of each variable against those figures.

1 *Working time*

Working time for both primary and secondary jobs is reported in 'buk3tk1' and 'buk3tk2' in IFLS1 (and in slightly different file names in the subsequent waves of IFLS). We use the labour time definition and measurement framework of the Food and Agriculture Organization (FAO, see Quiñones et al. 2009). There are two labour time dimensions: duration and frequency. Duration is the length of time that a job has continuously been performed by a specific person in a given time span, such as the number of months during the last year. The duration of a job can be from as short as one day to as long as one year. A full-year job is equal to 10 working months (42 weeks) or more per year, whereas a part-year job is less than 10 working months per year. Frequency, on the other hand, refers to how often a job is performed by an individual in a given time span, such as the number of hours per week. A full-time job is equal to 35 hours per week or more, while a part-time job is less than 35 hours per week (Quiñones et al. 2009: 3–4). Table A2 summarizes the results.

Table A2: Working time of respondents recorded in IFLS Book 3 (labour and employment)

Working time	**1993**		**1997**		**2000**		**2007**		**2014**	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
FYFT	5,485	56.2	6,652	55.6	9,308	53.8	11,451	56.8	13,191	53.9
PYFT	1,722	17.6	1,777	14.9	3,162	18.3	3,209	15.9	3,813	15.6
FYPT	1,473	15.1	2,176	18.2	3,048	17.6	3,544	17.6	4,794	19.6
PYPT	1,066	10.9	1,092	9.1	1,757	10.1	1,914	9.5	2,336	9.5
Total reporting sample	9,746	99.8	11,697	97.8	17,275	99.8	20,118	99.8	24,134	98.6
Total non-reporting sample	16	0.2	267	2.2	42	0.2	48	0.2	355	1.4
Population	9,762	100	11,964	100	17,317	100	20,166	100	24,489	100

Note: FYFT: full-year (≥42 weeks) full-time (≥35 hours); PYFT: part-year (<42 weeks) full-time (≥35 hours); FYPT: full-year >42 weeks) part-time (<35 hours); PYPT: part-year (<42 weeks) part-time (<35 hours).

Source: authors' construction.

2 *Occupation*

The types of occupations and their classification are based on the International Standard Classification of Occupations (ISCO-08), as stipulated by the International Labour Organization (ILO 2007). Table A3 summarizes the primary occupation of respondents.

Table A3: Primary occupation of respondents reported in IFLS Book 3 (labour and employment)

Primary occupation	1993		1997		2000		2007		2014	
	N	%	N	%	N	%	N	%	N	%
1	566	5.8	713	6.0	935	5.4	1,245	6.2	1,615	6.6
2	177	1.8	25	0.2	59	0.3	63	0.3	94	0.4
3	330	3.4	578	4.8	684	3.9	805	4.0	1,362	5.6
4	1,786	18.3	2,596	21.7	2,787	16.1	3,787	18.8	4,804	19.6
5	661	6.8	643	5.4	2,559	14.8	3,078	15.3	3,900	15.9
6	3,897	39.9	4,124	34.5	5,952	34.4	6,319	31.3	6,396	26.1
7	2,266	23.2	3,204	26.8	4,235	24.5	4,783	23.7	6,227	25.4
8	35	0.4	66	0.6	68	0.4	70	0.3	65	0.3
9	2	0.0	0	0.0	7	0.0	3	0.0	5	0.0
99	42	0.4	15	0.1	29	0.2	10	0.0	13	0.1
Total reporting sample	9,762	100	11,964	100	17,315	100	20,163	100	24,481	100
Total non-reporting sample	0	0.0	0.0	0.0	2	0.0	3	0.0	8	0.0
Population	9,762	100	11,964	100	17,317	100	20,166	100	24,489	100

Note: Primary occupation based on ISCO-08. 1: 'professional technical workers'; 2: 'administrative and managerial workers'; 3: 'clerical and related workers'; 4: 'sales workers'; 5: 'service workers'; 6: 'agricultural, animal husbandry, forestry workers, fisherman and hunters'; 7: 'production and related workers, transport operators, and labourers'; 8: 'military/police'; 9: 'students'; 99: 'unknown'.

Source: authors' construction.

3 Employment

There are eight categories of employment, summarized in Table A4: 'self-employed' (type 1); 'self-employed with household member assistant' (type 2); 'self-employed with permanent worker' (type 3); 'government worker' (type 4); 'private worker' (type 5); 'unpaid family worker' (type 6); 'casual worker in agriculture' (type 7); 'casual worker not in agriculture' (type 8); and 'unknown' (type 9).

Table A4: Employment type of respondents in IFLS Book 3 (labour and employment)

Employment type	1993		1997		2000		2007		2014	
	N	%	N	%	N	%	N	%	N	%
1	2,643	27.1	5,275	44.1	4,007	23.1	3,645	18.1	4,243	17.3
2	2,179	22.3	0	0.0	2,951	17.0	3,627	18.0	4,568	18.7
3	138	1.4	0	0.0	235	1.4	343	1.7	500	2.0
4	943	9.7	1,000	8.4	1,152	6.7	1,423	7.1	1,624	6.6
5	2,802	28.7	4,381	36.6	6,602	38.1	5,988	29.7	8,408	34.3
6	1,039	10.6	1,308	10.9	2,368	13.7	2,823	14.0	2,310	9.4
7	1	0.0	0	0.0	0	0	840	4.2	899	3.7
8	0	0.0	0	0.0	0	0	1,473	7.3	1,929	7.9
9	17	0.2	0	0.0	0	0	1	0.0	0	0.0
Total reporting sample	9,762	100	11,964	100	17,315	100.0	20,163	100.0	24,481	100.0
Total non-reporting sample	0	0	0	0	2	0.0	3	0.0	8	0.0
Population	9,762	100	11,964	100	17,317	100	20,166	100	24,489	100

Note: 1: 'self-employed'; 2: 'self-employed with household member assistant'; 3: 'self-employed with permanent worker'; 4: 'government worker'; 5: 'private worker'; 6: 'unpaid family worker'; 7: 'casual worker in agriculture'; 8: 'casual worker not in agriculture'; 9: 'unknown'.

Source: authors' construction.

4 *Sector*

Question 'tk19' in the IFLS specifies 10 sectors of the economy: 'agriculture, forestry, fishing, and hunting' (sector 1); 'mining and quarrying' (sector 2); 'manufacturing' (sector 3); 'electricity, gas, and water' (sector 4); 'construction' (sector 5); 'wholesale, retail, restaurants, and hotels' (sector 6); 'transportation, storage, and communications' (sector 7); 'finance, insurance, real estate, and business' (sector 8); 'social services' (sector 9); and 'other' (sector 10). However, there is no 'tk19' question in IFLS1. The results are presented in Table A5 below.

Table A5: Economic sector in which respondents work

Sectoral category of primary job	1993		1997		2000		2007		2014	
	N	%	N	%	N	%	N	%	N	%
1	0	0	4145	34.6	5,988	34.6	6351	31.5	6,543	26.7
2	0	0	78	0.7	94	0.5	125	0.6	299	1.2
3	0	0	1914	16.0	2,413	13.9	2671	13.2	3,143	12.8
4	0	0	56	0.5	53	0.3	59	0.3	121	0.5
5	0	0	651	5.4	743	4.3	926	4.6	1,149	4.7
6	0	0	2,637	22.0	3,835	22.1	4,977	24.7	6,166	25.2
7	0	0	509	4.3	679	3.9	672	3.3	570	2.3
8	0	0	92	0.8	118	0.7	165	0.8	1,124	4.6
9	0	0	1,867	15.6	3,359	19.4	4,188	20.8	5,022	20.5
10	0	0	15	0.1	26	0.2	29	0.1	336	1.4
Total reporting sample	0	0	11,964	100	17,308	99.9	20,163	100	2,4473	99.9
Total non-reporting sample	9,762	100	0	0	9	0.1	3	0	16	0.1
Population	9,762	100	11,964	100	17,317	100	20,166	100	24,489	100

Note: 1: 'agriculture, forestry, fishing, and hunting'; 2: mining and quarrying'; 3: 'manufacturing'; 4: 'electricity, gas, and water'; 5: 'construction'; 6: 'wholesale, retail, restraurants, and hotels'; 7: 'transportation, storage, and communications'; 8: 'finance, insurance, real estate, and business'; 9: 'social services'; and 10: 'other'.

Source: authors' construction.

Step 2: Defining and estimating the main variable (earnings)

The concept of earnings, as applied in wages statistics, relates to remuneration in cash and in kind paid to employees, as a rule at regular intervals, for time worked or work done, together with remuneration for time not worked, such as for annual vacation, other paid leave, or holidays. Earnings should include: direct wages and salaries, remuneration for time not worked (excluding severance and termination pay), bonuses and gratuities as well as housing and family allowances paid by the employer directly to employees (Quiñones et al. 2009). In the IFLS, earnings data are recorded in Book 3 section TK, which distinguishes various types of earnings that are related to different types of work and are thus within the scope of the abovementioned concept of earnings. The IFLS also has sufficient information on time worked, although some responses are reported in weekly intervals, others monthly or annually. In this research all time units are converted to monthly.

Income data are extracted from each type of employment, in both primary and secondary jobs. Each type of employment corresponds to a category of personal income sources and relates to a specific question in the IFLS. For instance, employment types 1 to 3 usually answer a question such as 'Approximately how much net profit did you gain last month (variable tk26a1)/year (tk26a3)?'. A complete list of IFLS questions and their related type of employment is provided in Table A6.

Table A6: Earnings concept and IFSL questionnaire for each type of employment

Earnings concept	IFLS questionnaire	Employment type							
		1	2	3	4	5	6	7	8
Profit	Approximately how much net profit did you gain last month (TK26A1)/year (TK26A3), after subtracting all your business expenses?	√	√	√					
Bonuses/ gratuities	What is the amount of year-end-bonus or other bonuses you received during the last year (TK25A2b)?				√	√		√	√
Direct wages/ salaries	Approximately what was your salary/wage during the last month (including the value of all benefits) (TK25A1)?				√	√		√	√

Note: 1: 'self-employed'; 2: 'self-employed with household member assistant'; 3: 'self-employed with permanent worker'; 4: 'government worker'; 5: 'private worker'; 6: 'unpaid family worker'; 7: 'casual worker in agriculture'; and 8: 'casual worker not in agriculture'.

Source: authors' construction.

As expected, some respondents did not report their earnings. However, three sources of earnings are not sufficient to be used as covariates to impute a non-reporting value of bonuses and profit. For instance, year-end bonuses would probably relate to worker performance during the year, which requires information we do not have. Profit might be related to goods/services produced/sold, for which we again do not have sufficient information. Fortunately, there is the possibility to impute non-reported values of salaries since they are strongly related to several covariates available. These covariates are: (1) type of occupation, (2) type of employment, (3) economic sector, (4) working time, and (5) location where the respondent worked. We assume that workers with similar covariates have similar salaries.

Salary data are imputed in only two cases: (1) where salary is non-reported but there is complete information on covariates and (2) in the case of outliers. Based on the abovementioned assumption, we calculate median and standard deviation values at provincial level for the same group of covariates. We define outliers as all values higher than the median plus three times the standard deviation, or all values lower than the median minus three times the standard deviation. By taking this moderate approach to the imputation process, more than 99 per cent of our data is the original raw data drawn from the survey, except for 1993 (97.5 per cent). A summary of all salary data from primary jobs, before and after imputation, is presented in Table A7.

Table A7: Primary job wage/salary: numbers of reported data

IFLS year	1993	1997	2000	2007	2014
Before imputation (n)	3,592	5,252	7,574	9,461	12,545
After imputation (n)	3,679	5,266	7,596	9,529	12,588
Imputed values (%)	2.42	0.27	0.29	0.72	0.34

Source: authors' construction.

As with primary jobs, data on secondary jobs are available in almost all cases and are thus taken from the IFLS, with a few caveats. During the imputation process we found that four observations in 1997 and one in 2014 do not have sufficient data on covariates, while the earnings data for all four observations marked them as outliers. As our imputation process dictates, we did not impute any observation with outlier values and incomplete covariates. A summary of labour wage data for secondary jobs is provided in Table A8.

Table A8: Secondary job wage/salary: numbers of reported data

IFLS year	1993	1997	2000	2007	2014
Before imputation (n)	575	506	1,052	1,214	1,843
After imputation (n)	588	502	1,056	1,222	1,842
Imputed values (%)	2.26	-0.79	0.38	0.66	-0.05

Source: authors' construction.

The imputation process of secondary-job salaries shows similar results to that of primary-job salaries, where we can still rely on original data (before imputation) taken from raw data of the survey (at around 99 per cent except for 1993 (97.74 per cent)). The next step in extracting individual earnings data is to combine the total of individual labour income or earnings from all sources (profit, salary, and bonus) from primary jobs, secondary jobs, and both. The primary sources of earnings for primary jobs are summarized in the Table A9.

Table A9: Reported primary sources of earnings for primary jobs: broken down by type of employment

Earnings source	Year	Employment type of primary job									Total
		1	2	3	4	5	6	7	8	99	
Profit	1993	1,869	1,440	103	0	0	0	0	0	0	3,412
	1997	4,962	0	0	0	0	0	0	0	0	4,962
	2000	3,838	2,832	232	0	0	0	0	0	0	6,902
	2007	3,479	3,484	333	0	0	0	0	0	0	7,296
	2014	4,044	4,352	473	0	0	0	0	0	0	8,869
Salary	1993	0	0	0	943	2,732	0	0	0	4	3,679
	1997	0	0	0	989	4,277	0	0	0	0	5,266
	2000	0	0	0	1,149	6,447	0	0	0	0	7,596
	2007	0	0	0	1,421	5,941	0	753	1,414	0	9,529
	2014	0	0	0	1,604	8,307	0	820	1,857	0	12,588
Bonus	2007	0	0	0	1,241	4,514	0	199	578	0	6,532
	2014	0	0	0	1,257	6,189	0	174	657	0	8,277

Note: 1: 'self-employed'; 2: 'self-employed with household member assistant'; 3: 'self-employed with permanent worker'; 4: 'government worker'; 5: 'private worker'; 6: 'unpaid family worker'; 7: 'casual worker in agriculture'; 8: 'casual worker not in agriculture'; 99: 'unknown'. A question on bonuses/gratuities is only asked in the two latest editions of IFLS.

Source: authors' construction.

The pattern uncovered in Table A9 is consistent with that discussed in Table A6, namely that sources of earnings (profit, salary, and bonus) are correlated with type of employment. For instance, employment type 4 (government workers) or type 5 (private workers) are those who have salary and bonus as their income sources, but they of course do not have profit as their earnings source. A similar pattern is also evident for secondary jobs, which are summarized in Table A10.

Table A10: Reported secondary sources of earnings for secondary jobs: broken-down by type of employment

Earnings sources	Year	Employment type of secondary job									Total
		1	2	3	4	5	6	7	8	99	
Profit	1993	529	335	16	0	0	0	0	0	0	880
	1997	1,008	0	0	0	0	0	0	0	0	1,008
	2000	1,287	839	77	0	0	0	0	0	0	2,203
	2007	1,082	1,273	94	0	0	0	0	0	0	2,449
	2014	1,644	1,446	154	0	0	0	0	0	0	3,244
Salary	1993	0	0	0	28	554	0	0	0	6	588
	1997	0	0	0	16	486	0	0	0	0	502
	2000	0	0	0	46	1,009	0	0	0	1	1,056
	2007	0	0	0	60	391	0	369	402	0	1,222
	2014	0	0	0	76	766	0	347	653	0	1,842
Bonus	2007	0	0	0	40	226	0	181	179	0	626
	2014	0	0	0	52	351	0	108	183	0	694

Note: Questionnaire on bonuses/gratuities is only asked in the two latest editions of IFLS. 1: 'self-employed'; 2: 'self-employed with household member assistant'; 3: 'self-employed with permanent worker'; 4: 'government worker'; 5: 'private worker'; 6: 'unpaid family worker'; 7: 'casual worker in agriculture'; 8: 'casual worker not in agriculture'; 99: 'unknown'.

Source: authors' construction.

Finally, having extracted all earnings data from primary and secondary jobs, we sum total monthly individual labour incomes or earnings. These total earnings can be detailed on the basis of the covariates used in imputation. Table A11 shows an example of earnings broken down by employment type.

Table A11: Reported total earnings from all sources

Employment type	earning93		earning97		earning00		earning07		earning14	
	N	%	N	%	N	%	N	%	N	%
1	1,935	20	4,387	37	3,846	22	3,506	17	4,040	16
2	1,500	15	-	-	2,821	16	3,478	17	4,306	18
3	104	1	-	-	228	1	333	2	478	2
4	939	10	982	8	1,139	7	1,408	7	1,584	6
5	2,773	28	4,134	35	6,234	36	5,816	29	8,141	33
6	-	-	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	739	4	770	3
8	-	-	-	-	-	-	1,369	7	1,789	7
99	4	0	-	-	-	-	-	-	-	-
TRS	7,255	74.32	9,503	79.43	14,268	82.39	16,649	82.56	21,108	86.19
TNRS	2,507	25.68	2,461	20.57	3,049	17.61	3,517	17.44	3,381	13.81
Population	9,762	100	11,964	100	17,317	100	20,166	100	24,489	100

Note: 1: 'self-employed'; 2: 'self-employed with household member assistant'; 3: 'self-employed with permanent worker'; 4: 'government worker'; 5: 'private worker'; 6: 'unpaid family worker'; 7: 'casual worker in agriculture'; 8: 'casual worker not in agriculture'; 99: 'unknown'. TRS = Total reporting sample; TNRS = Total non-reporting sample. A question on bonuses/gratuities is only asked in the two latest edition of IFLS.

Source: authors' construction.

Step 3: Creating temporal and spatial deflators

1 Temporal deflator

Earnings is a variable with monetary value, which is recorded in nominal terms in the IFLS. However, as the IFLS has been conducted through a 21-year period (1993–2014), it is necessary to standardize the monetary variable so that we can compare the true value of the variable across time. We call this standard to be used to transform the nominal values a *temporal deflator*. We choose 2014 as the base year for the temporal deflator because it enables an easy interpretation of the direction and meaning of the change in monetary variables, such as earnings, over the years. That is, we expect the real earnings of individuals to increase from 1993 over the following years (1997, etc.) unless there are some exceptional economic incidents such as economic crises at a particular point in time.

To create a temporal deflator and thus to standardize monetary values to a selected base year, we need consumer price index (CPI) data for each location unit (usually cities) where the IFLS was conducted. Generally speaking, there are two options. The first is to create our own CPI based on the data available (prices and quantity data). The second is to use readily available CPI data taken from the IFLS location or at least from similar/neighbouring locations.

For the first option, we need price and quantity data of goods and services collected by the IFLS along with household consumption expenditure data. Unfortunately, these are not available. While some price information was collected in the household questionnaire and separately for local markets in the community questionnaire, there is only a limited number of commodities available (Strauss et al. 2004; Witoelar 2009). Another source of price and quantity data can be used to create a CPI, namely Indonesia's National Socioeconomic Survey (SUSENAS). However, unlike the IFLS, which is publicly available, SUSENAS microdata are not, but can only be purchased at great expense via Badan Pusat Statistik (BPS). In short, the first option for creating our own CPI is not viable in this research.

Turning to the second option, there are researchers who have created a CPI using SUSENAS such as Friedman and Levinsohn (2001), Strauss et al. (2004), and Witoelar (2009). However, their works are not publicly available and, even if they were, they would not be sufficient because they do not include data from the year 2014. This leaves us with the CPI officially published by BPS. BPS has been conducting CPI surveys since 1979 with several changes of names, base years, and survey location (cities). All those changes are summarized in Table A12.

Table A12: Changes in BPS's CPI surveys since 1979

Period	Index name	Base month and year	Number of provincial capital cities surveyed
Before April 1979	Cost of Living Index	September 1966	n.a.
April 1979 – March 1990	Consumer Price Index	April 1977 – March 1978	17
April 1990 – November 1997	Consumer Price Index	1988/1989	27
December 1997 – December 2003	Consumer Price Index	1996	44
January 2004 – May 2008	Consumer Price Index	2002	45
June 2008 – December 2013	Consumer Price Index	2007	66
Since January 2014	Consumer Price Index	2012	82

Source: authors' elaboration based on BPS.

Those changes and reconciliations are published in PDF files archived in the BPS website, searchable and downloadable from this page: https://www.bps.go.id/pencarian.html. We found that searching using the term 'Indikator Ekonomi' yielded the data we were looking for. The earliest publication available is 'Indikator Ekonomi Januari 1999'. Despite being unpractical, the publication is worthwhile because it not only provides a detailed and long time span but also offers CPI data and their changes over time (inflation rate). Moreover, 'Indikator Ekonomi Januari 1999' contains data on the inflation rate of 44 cities in 1994.

After a repeated process of searching and downloading 'Indikator Ekonomi Januari 2000', 'Indikator Ekonomi Januari 2001', etc., the next step was to manually enter inflation rate data for the surveyed cities, 44 in 1997, 45 in 2004, 66 in 2008, and 82 in 2014. Finally, we reconciled all the inflation rate data retrieved from the 'Indikator Ekonomi' files of various years. This resulted in an incomplete file with missing values for some years. We replaced missing values with those of neighbouring cities. In short, we now had a complete list of inflation rates for 82 cities in Indonesia from 1993 to 2014.

The next step was to merge these data with all cities that appeared in the IFLS, which amounts to 326 cities in any of the five waves. Only 62 of these cities are also included in BPS's CPI surveys. For the remaining 264 cities, we assigned the provincial average inflation rate value. If this was also missing, we assigned the value of neighbouring cities. In the end, all 326 IFLS cities were assigned local inflation rate (*loci*) values for the period 1993 to 2014.

Subsequently, we created a temporal deflator by using 2014 as a base year. This means that we set the value of the 2014 deflator for all 326 IFLS cities as 1 (base year). For the year 2013 and all other years until 1993, we used this formula to calculate the temporal deflator and applied it backtracking from 2013 to 2012, etc.:

$$\text{tdef\_b14\_YN} = \text{tdef\_b14\_Yn} / (\text{tdef\_b14\_Yn} + \text{loci\_YN}/100) \qquad (3)$$

In this formula, *tdef_b14_YN* represents the temporal deflator base year 2014 for year n, *tdef_b14_Yn* refers to the temporal deflator base year 2014 for year n+1, and *loci_YN* represents the inflation rate of the particular city in year n.

Now we turn to the second deflator, which allows us to standardize the spatial differences in our sample.

2 Spatial deflator

The basic idea of a spatial deflator is to determine the difference in what the same amount of money can buy in different places (province, city, urban, rural, etc). The commonly used standard for this purpose is the poverty line of each location standardized against the poverty line of a particular spatial reference. Poverty lines are appropriate as they are usually based on food expenditures necessary for nutritional adequacy and some allowance for 'essential' non-food items (Pradhan et al. 2000). This means that poverty lines reflect the amount of money needed to purchase the goods and services required (standardized) to live a decent life in a particular location.

To that end, ideally, we use the poverty line of each district in the IFLS and select one particular district as a point of reference. However, there are no readily available data on this, while constructing our own poverty line would require a lot of effort which is outside the scope of this research. The second-best method is to use poverty lines at provincial level disaggregated by urban and rural areas in each province. There are studies that have used this approach, such as Pradhan et al. (2000), and Strauss et al. (2004). However, we were not able to use their data because we needed poverty lines from 1993 to 2014. We therefore turned to BPS's online data source and the Indonesia Database for Policy and Economic Research (DAPOER) to collect a list of poverty lines at provincial level but not disaggregated by urban–rural areas. We decided to use the poverty line of DKI Jakarta as the base, which means that we divided all poverty lines by that of DKI Jakarta. Thus, we used the following formula:

$$\text{sdef\_bDKI\_YN}=\text{plineYN}/\text{plineDKI\_YN} \tag{4}$$

where *sdef_bDKI_YN* represents the spatial deflator base DKI Jakarta for any particular year (YN), *plineYN* denotes the poverty line of any province in any particular year, and *plineDKI_YN* refers to the poverty line of DKI Jakarta in that particular year. Applying the above calculation to all provinces in the IFLS resulted in a spatial deflator. Subsequently, we combined the temporal and spatial deflators to produce our final deflator.

3 Final deflator: Temporal and spatial deflator combined

To create our final deflator, a simple formula was applied:

$$\text{def\_ YN}= \text{tdef\_b14\_YN}/\text{sdef\_bDKI\_YN} \tag{5}$$

where *def_YN* refers to the temporal (2014) and spatial deflator (DKI Jakarta) combined for any particular year (YN), as a result of the temporal deflator being divided by the spatial one. Having created the deflator, we subsequently applied it to the nominal earnings values in our data as produced in Step 2. Table A13 compares the number of observations and the mean and median of nominal and real earnings in Indonesian Rupiah (IDR) and US Dollars (USD).

Table A13: Number of observations, mean, and median: nominal vs real earnings (IDR, 2014)

Year		1993	1997	2000	2007	2014
N		7,255	9,503	14,268	16,649	21,108
Mean Earnings	Nominal (IDR)	348,463	235,865	466,161	10,145,333	49,734,764
	Real (IDR)	289,390	144,531	138,872	1,277,403	4,123,780
	Real (USD)	NA	NA	14.47	135.6	331.5
Median Earnings	Nominal (IDR)	82,000	135,000	250,000	600,000	1,216,667
	Real (IDR)	69,564	78,852	72,900	70,417	105,866
	Real (USD)	NA	NA	7.6	7.48	8.51

Notes: Real values are nominal values deflated by spatial (DKI Jakarta as base value) and temporal (2014 as base value) deflators. Earnings are calculated as a monthly value. The applied exchange rate of IDR to USD, retrieved from BPS's publication 'Selected foreign exchange middle rates against rupiah at Bank of Indonesia and prices of gold in Jakarta', accessed 4 July 2020, is 1 USD = 9,595 IDR (2000), 9,419 IDR (2007), and 12,440 (2014).

Source: authors' construction.