



UNITED NATIONS
UNIVERSITY
UNU-WIDER

WIDER Working Paper 2018/148

Income under-reporting and tax evasion

How they impact inequality in Vietnam

Hai Anh La¹ and Duc Anh Dang²

December 2018

Abstract: Personal income tax is attracting more attention from the Vietnamese government, which has been looking for a way to reinforce its budget revenue. Although this tax plays an increasing role, representing 7.3 per cent of the revenue expected in 2018, this figure is still small, suggesting an issue of tax evasion and ineffective tax policy. Using the Vietnam Household Living Standard Surveys 2010, 2012, 2014, and 2016 and the expenditure-based approach pioneered by Pissarides and Weber, this paper first applies the non-linear least squares method to distinguish under-declaration rates for various income sources, and then uses a static microsimulation SOUTHMOD model to estimate the impact of income under-reporting on the scale of tax evasion and income inequality of Vietnam. The paper finds that the officially reported income only accounts for 80 per cent of the true income, leaving 20 per cent unreported. Consequently, without income under-reporting, tax revenue in Vietnam would increase by about VND23,000 billion (equivalent to US\$1.03 billion) and the Gini coefficient for disposable income would increase from 0.379 to 0.409.

Keywords: Engel curve, income inequality, income under-reporting, permanent income, tax evasion, Vietnam

JEL classification: D12, H24, H25, H26

Acknowledgements: The results presented here are based on VNMOD v1.0. VNMOD is developed, maintained and managed by UNU-WIDER in collaboration with the EUROMOD team at ISER (University of Essex), SASPRI (Southern African Social Policy Research Insights) and local partners in selected developing countries (Ethiopia, Ghana, Mozambique, Tanzania, Zambia, Ecuador, and Vietnam) in the scope of the SOUTHMOD project. The local partner for VNMOD is Central Institute for Economic Management (CIEM). We are indebted to the many people who have contributed to the development of SOUTHMOD and VNMOD. The results and their interpretation presented in this publication are solely the authors' responsibility.

We gratefully acknowledge the comments and suggestions made by the participants of the Tax-Benefit Microsimulation Analysis Workshop organized on 14 June 2018 in Helsinki, Finland. We especially would like to thank Nguyen Anh Duong at CIEM for his comments and support related to the SOUTHMOD model.

¹ National Centre for Social and Economic Modelling, University of Canberra, Canberra, Australia; ² National Centre for Socio-Economic Information and Forecasting, Ministry of Planning and Investment, Ha Noi, Viet Nam, corresponding author: haianhla2011@gmail.com

This study has been prepared within the UNU-WIDER project on 'SOUTHMOD – Simulating Tax and Benefit Policies for Development' which is part of the Institute's larger research project on 'The economics and politics of taxation and social protection'.

Copyright © UNU-WIDER 2018

Information and requests: publications@wider.unu.edu

ISSN 1798-7237 ISBN 978-92-9256-590-9 <https://doi.org/10.35188/UNU-WIDER/2018/590-9>

Typescript prepared by Gary Smith.

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

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

Katajanokanlaituri 6 B, 00160 Helsinki, Finland

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

1 Introduction

Recent years have seen an increase in studies measuring to what extent income collected from household surveys is under-reported. People tend to distort their actual income when reporting to administrative authorities if their incentives for doing so are sufficiently large (for example, to evade tax payments) or the cost of doing so is relatively small (such as minimal sanctions on tax evasion and fraud), or both. In addition, it tends to be more comfortable for individuals who intend to misreport their income to tax authorities to re-use the same report for household surveys, and these individuals may feel obliged and face few costs in providing consistent measures to both tax authorities and survey data collectors (see Hurst and Pugsley 2010).

Several researchers have tried to estimate the extent of under-reporting by using the information on reported income and consumption and the expenditure-based approach pioneered by Pissarides and Weber (1989). The fundamental idea is that all interviewees accurately report their expenditure, or there is no systematic difference in misreporting consumption and income. One type of employment (wage and salary workers in most cases) is generally assumed to provide a systematically unbiased report of income to household surveys. Other sources of income (for example, self-employed income) tend to be under-reported, and the extent of this under-reporting is estimated using excess food consumption, given an observed income level. In most cases, families with farm income are removed from the samples.

Pissarides and Weber's (1989) approach has been applied in many studies estimating tax evasion and the size of the black economy. Various high- or upper-middle-income countries have been examined, including Britain in 1982 (Pissarides and Weber 1989) and 1993 (Lyssiotou et al. 2004), Estonia during 2002–07 (Kukk and Staehr 2014, 2017), Finland in 1994–96 (Johansson 2005), North Cyprus in 1999 (Besim and Jenkins 2005), South Korea during 2000–05 (Kim et al. 2017), Russia during 2002–05 (Kim et al. 2017), Spain in 2006–09 (Martínez-Lopez 2013), Sweden over the periods 1999–2001 and 2003–04 (Engström and Holmlund 2009) and during 2003–09 (Engström and Hagen 2017), Turkey in 2004–05 (Davutyan 2008), and the USA in 1999 (Feldman and Slemrod 2007) and over the periods 1980–2003 and 1980–97 (Hurst and Pugsley 2010).

Results from these empirical studies commonly support the hypothesis that self-employed individuals tend to under-report their incomes. However, the scale of this under-reporting varies considerably across countries. On average, it ranges from 36 per cent of true income for British self-employed households in 1982, 54 per cent and 39 per cent for those with the household head in blue-collar and white-collar occupations in 1993, respectively; to 57–62 per cent in Estonia; some 16–40 per cent in Finland; 29 per cent in both Korea and Russia; 25 per cent for the Spanish self-employed to obtain the level reported by employees; 10–14 per cent for both the self-employed and those working for private companies in North Cyprus; around 22–30 per cent in Sweden; 20 per cent in Turkey; 35 per cent, 78 per cent, and 74 per cent for American households reporting positive self-employed, non-farm small business, and farm income, respectively in 1999, but only 25 per cent for the USA's self-employed families over the period 1980–2003.

While studies on income under-reporting for tax purposes frequently focus on high-income countries of which revenues from personal income tax (PIT) are already proved to play a moderate role in total tax revenues, it is difficult to find similar research on a developing country where the

PIT share of total tax revenue tends to be much smaller,¹ possibly implying more severe tax evasion problems. As one of the developing countries, Viet Nam is, therefore, an interesting case to study. In this country, the PIT is attracting more attention from the central government, which has been looking for a way to reinforce the budget revenue and avoid the risk of an enormous budget deficit through reverting to domestic revenues as a replacement for trade revenues.² According to the Ministry of Finance of Viet Nam, in 2008, the PIT accounted for only 3 per cent of total tax revenue, while taxes from crude oil and imported/exported goods and services contributed 42 per cent. Since 2009, when tax reforms were implemented and the PIT started to be applied to a broader group of income earners, personal tax has played an increasingly important role. In 2016, the shares of total tax revenue increased moderately to 5.9 per cent for this tax source, while reducing sharply to 3.6 per cent for crude oil and 15.5 per cent for imported/exported goods and services. The contribution of the PIT to the overall tax revenue is expected to continue rising, to reach 7.3 per cent in 2018. However, compared to developed countries, this figure is still small, suggesting more serious issues of tax evasion and ineffective tax policy.

Unlike developed countries, Viet Nam is characterized by a dominant share of the population living in rural areas (70 per cent), leading to the primary income source in many households originated from farm activities. It is, therefore, hard to exclude farmers in Vietnamese studies. This income, together with private remittances (both from domestic and international sources), pension payments, and other social benefits, are tax-free. Hence, instead of assuming wages and salaries as accurately recorded earnings, as in previous studies, for the case of Viet Nam we consider the tax-free income sources to be a correctly reported component. We also incorporate wages and salaries in foreign-invested enterprises (FIEs), despite their taxable income characteristics, as a part of this component. This is because, compared to state-owned enterprises (SOEs) and private firms, workers in FIEs tend to have higher official salaries and generally do not have much time left to obtain supplemental income, resulting in a reduced likelihood of income under-reporting. Additionally, tax collusion and corruption tend to be lower in this sector due to the better accounting records following the Vietnamese regulations. As a result, our unbiased recorded income component includes not only tax-free income sources but also wages or salaries from the foreign sector.

In contrast, both non-farm self-employment income and wages or salaries in public and private sectors are taxable and more likely to be misreported. An indicator of tax under-declaration and non-compliance among non-farm household businesses in Viet Nam is that the current PIT system relies massively on wage and salary earners (their tax payment accounted for 74 per cent of total PIT revenues in 2011, according to Nguyen and Liu (2014)) but does not reflect the actual labour market situation (30.4 per cent of all workers received wages or salaries in 2011 (see OECD 2014)). It is mainly due to the weak capacity of authorities in collecting taxes and reaching household businesses, especially when cash is widely used in transactions and the size of the informal sector is significant.

Income under-reporting can also happen among Vietnamese wage and salary earners, both in the public and private sectors. Salaries in the public sector tend to be low and depressed, driving employees to seek extra income (Painter 2003), which is more likely to be misreported. Additionally, many firms in the formal sector under-declare their staff wages to avoid paying full

¹ According to Bird and Zolt (2014), in the same period, while PIT explained about 25 per cent of total tax revenues (over 7 per cent of gross domestic product (GDP)), on average for high-income nations this number was only 9 per cent (less than 2 per cent of GDP) for developing countries.

² When integrating more deeply into the global economy since 2007, Viet Nam had to cut several tariffs following free trade agreements signed with ASEAN and other countries, leading to a significant drop in tax revenues.

social insurance contributions. Workers also benefit from this practice as enterprises paying lower contributions tend to pay higher net wages to them under the counter (Castel and To 2012). Moreover, corruption is another strong reason for explaining income misreporting in these sectors. With the systemic feature of street-level bureaucracy (Kerkvliet 1995), corruption in Viet Nam is severe. According to the 2017 Corruption Perception Index from Transparency International, Viet Nam ranks 107 out of 180 countries and scores 35 on a scale from 0 (highly corrupt) to 100 (very clean). Using two rounds of small- and medium-sized enterprise surveys in 2005 and 2007, Rand and Tarp (2012) find that larger, more formal, more profitable firms, those with higher capital–labour ratios and more interactions with the government, are more likely to pay bribes. With there being almost no harsh penalties for tax evasion and minimal government monitoring, salaried workers in both the public and private sectors have plenty of opportunities to hide their actual income from the authorities, which could result in substantial tax revenue losses. Investigating the scale of tax evasion as well as how income under-reporting and tax evasion impact inequality is, therefore, crucial to Vietnamese policy makers.

Our goals in this paper are twofold. First, we examine the extent to which non-farm self-employed individuals and wage or salary workers in public and private sectors under-declare their income, using the Viet Nam Household Living Standard (VHLSS) surveys for 2010, 2012, 2014, and 2016. The second goal of the paper is to employ the static microsimulation SOUTHMOD model for the case of Viet Nam to calculate the PIT that should be paid if there is no income under-reporting. From this calculation, we can estimate the scale of tax evasion and measure its effects on inequality.

This paper makes a number of contributions to the current literature. First, it is the first study on income under-reporting in a developing country that is dominated by farming activities. Second, we adopt the methodology of Lyssiotou et al. (2004) and Feldman and Slemrod (2007), extended from Pissarides and Weber's (1989) analysis framework, to capture the multiple-source income characteristics more accurately and cover all Vietnamese households rather than narrowing down to sub-samples, as is usually observed in previous studies.³ Including the whole population and being flexible in income components mitigates the under- or over-estimation issues of tax evasion and allow us to measure the inequality index for the whole population. Third, the paper estimates not only income under-declaration but also its impact on income inequality using the microsimulation model, which allows for the family behavioural changes towards minimizing the total PIT paid to the authorities. Currently, very few papers examine the impact of tax evasion on income inequality, although it is commonly recognized that tax evasion can cause severe inequities and inefficiencies (see Slemrod 2007). To measure inequality, we calculate not only the Gini index as an overall measurement, but also the income ratios at different segments of the spectrum to capture a broader picture of income inequality impact in Viet Nam.

The paper is organized as follows. Section 2 provides an overview of the current PIT law in Viet Nam. Section 3 discusses the methodology. Section 4 describes the data and variables. The results are presented in Section 5. Section 6 presents the conclusions.

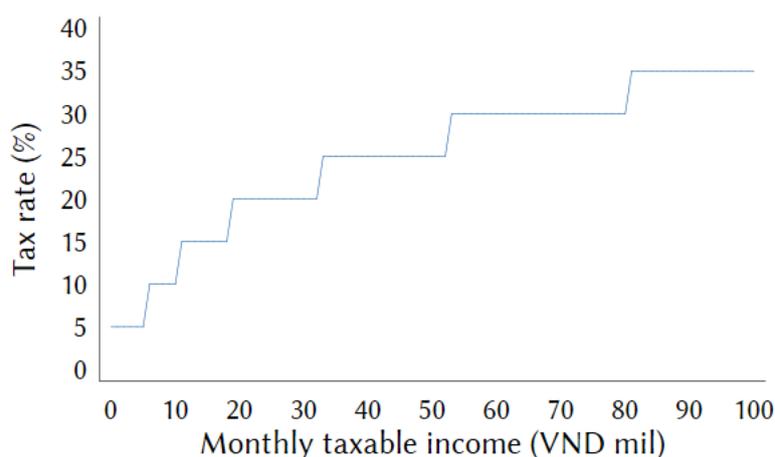
³ In developed countries, family income principally comes from one or two sources: wages of non-farm self-employment income, so restricting to sub-samples with either one or the other source of income without any farm earnings is possible. However, in Viet Nam households tend to rely on various income sources, including not only wages and non-farm self-employment income, but also farming income and remittances.

2 Personal income tax in Viet Nam: an overview

At the end of 1986, Viet Nam decided to move from a centrally planned mechanism to a market-oriented economy, followed by a series of economic renovations. Taxation reform started in the 1990s when the National Assembly passed important tax laws to be applied uniformly to several economic sectors. Among various tax reforms, the PIT law was passed by the National Assembly in November 2007 and went into effect in January 2009. Before that, the taxation of personal income was regulated by Ordinance No. 35/2001/PL-UBTVQH10 on high-income earners only. Like an excise tax, this ordinance covered very few taxpayers (0.3 per cent of the total population in 2003 (see Shukla et al. 2014)). The high exemption level (VND5 million per month) excluded the majority of Vietnamese people from the tax base and allowed a significant income tax contribution from foreigners (62.4 per cent in 2003 (see Shukla et al. 2014)).

The PIT tax law (Law No. 04/2007/QH12) on wages or salaries is characterized by a progressive system with seven tax brackets at 5, 10, 15, 20, 25, 30, and 35 per cent (see Figure 1). This law grants tax declaration and payment carried out on a withholding basis. Although individuals can declare their own taxes, the tax authorities usually require employers to collect taxes on behalf of their employees. Therefore, the employers must withhold a given proportion of their employees' income equivalent to their respective tax liabilities and deposit the withheld amounts in the state treasury. At the end of the year, the employers will finalize the PIT on behalf of their employees if their employees have income from them only or any irregular income from other sources that does not exceed VND10 million per month. Otherwise, they are responsible for self-declaration and payment of tax. Tax refunds due to excess tax payments are available in Viet Nam. Under this law, in addition to misreporting irregular income to employers or under-declaring gross income directly to tax authorities, there is also potential collusion between high- and low-income earners to lower their tax payments. Low-income earner can be assigned an extra income which will later be transferred to the high-income earners, on the condition that they can benefit from tax refunds for this additional income at the end of the financial year. At the same time, the high-income earners under-report their incomes to pay less tax.

Figure 1: Personal employment income tax rates since 2009



Source: Authors', based on Decree No. 100/2008/ND-CP.

Employment income (wages and salaries) include taxable and non-taxable parts. The non-taxable portion includes allowances such as those for security, survivorship or hazard purposes, labour accidents or working capacity loss, maternity leave, retirement, severance and job loss, and other allowances paid by the social insurance system. Meanwhile, contributions to mandatory social,

health, and unemployment insurance schemes,⁴ as well as contributions to voluntary pension schemes and for charities, are deducted from the taxable component. Supplementary deductions include VND9 million/month for any income earner and VND3.6 million/month⁵ per dependant of the taxpayer based on their declaration, according to revisions of the tax law (Law No. 71/2014/QH13 and Circular No. 92/2015/TT-BTC). A qualifying dependant can be: a child aged under 18; an adult child studying for a Bachelor degree or lower; a child, a spouse, or a parent who is incapable of working; or a parent beyond working age. They should all have no income or income not exceeding VND1 million/month. Furthermore, there is no limit to the number of dependants in declaring tax deductions, but each dependent can only be attributed to one taxpayer in a family.

Regarding non-employment income, the taxable part consists of business revenue (including rental income) exceeding VND100 million/year; investment income (e.g. interest, dividends); gains on the sale of shares or real estate; and inheritances over VND10 million. This component excludes interest earned on deposits with credit institutions and on life insurance policies; compensation paid under insurance policies; retirement pensions; international remittances; and income from the transfer of properties or inheritances/gifts between family members. In addition, income from agricultural products that have not been processed or have only been through preliminary processing is exempt from PIT payment. Before 2015, taxable non-employment income was taxed under the same progressive tax system of employment income. From 2015, self-employed individuals have to pay PIT at the deemed rates on the revenue from the sale of goods and provision of services if their business revenue exceeds VND100 million per year. The tax rate of 0.5 per cent is applied for revenue from wholesale or retail activities; 1.5 per cent for mining activities, manufacturing, transportation, services with goods provision, construction services with material provision; 2 per cent for services and construction (except for material/goods provision); 5 per cent for asset leasing; and 1 per cent for supply and distribution of other goods and activities. In terms of the non-business income, the deemed rates are 5 per cent on interest/dividends, 0.1 per cent on sales of shares, 2 per cent on sales of real estate, and 10 per cent on inheritance/gifts.

The regulations also set out penalties for tax offences. According to Circular No. 166/2013/TT-BTC, the fine for under-declaring tax payable or over-declaring tax exemption or tax deduction is 20 per cent of the tax arrears or refundable tax. In the case of taxpayers not declaring tax, the penalty is 1.5, 2, or 3 times the tax arrears, depending on the determined level of violation. Meanwhile, late payment of tax is subject to an interest rate of 0.05 per cent of the tax liability per day late. Thus, there are much more harsh punishments for tax avoidance, suggesting the more popular strategy of income under-reporting compared to the decision not to declare tax.

⁴ The contribution rates for employees are 8 per cent of their wages or salaries for social insurance, 1.5 per cent for health insurance, and 1 per cent for unemployment insurance. However, these contributions are capped at 20 times the minimum salary (VND1.390 million/month since July 2018) for social and health insurance and 20 times the minimum regional salary for unemployment insurance. The minimum regional wages range from VND2.76 million/month for region IV to VND3.09 million/month for region III, VND3.53 million/month for region II, and VND3.98 million/month for region I. Meanwhile, employers pay the rates at 18 per cent, 3 per cent, and 1 per cent of their workers' wages or salaries to further contribute to the social, health, and unemployment insurances of these employees.

⁵ These figures were VND4 million/month and VND1.6 million/month in 2009, respectively.

3 Methodology

In this section, we first present the method to estimate income under-reporting for tax purposes in the case of Viet Nam and then discuss the microsimulation model to measure the impact of tax non-compliance on income inequality.

3.1 Income misreporting measurement: a consumption-based approach

Our empirical strategy is similar to the one outlined by Pissarides and Weber (1989). However, rather than differentiating between self-employment and other-employment income, we assume that a tax-free component including farm self-employment income, remittances, pension payments, and other social benefits, together with earnings from FIEs, are correctly reported. Then we estimate the under-reporting of income portions from non-farm self-employment revenue, wage or salary in the public and private sectors. Because a characteristic of Vietnamese households is having income from various sources, we apply the non-linear least squares method proposed by Lyssiotou et al. (2004) and Feldman and Slemrod (2007) to distinguish the under-declaration rates among these sources. We also adopt the Engel curve to describe the food expenditure and income relationship conditional on household characteristics.

We assume that the gross true income y_{it}^T of household i at time t is:

$$y_{it}^T = \sum_{k=0}^6 \delta_k y_{ikt} \quad (1)$$

where y_{ikt} is gross reported income from source k ; k can be tax-free components and wages from FIEs, which is denoted by y_{i0t} , unregistered or registered non-farm self-employment income, wages/salaries in other household businesses, public and private firms, rental and investment income; δ_k is a factor measuring the extent of under-reporting income from source k and equals 1 if k refers to tax-free components and wages from FIEs (i.e. $\delta_0 = 1$) and is expected to be greater than 1 otherwise.

Equation 1 can be rewritten as:

$$y_{it}^T = y_{it} (w_{i0t} + \sum_{k=1}^6 \delta_k w_{ikt}) \quad (2)$$

where $w_{ikt} = y_{ikt} / y_{it}$ is the share of gross reported income from source k , except for farm-employment income, which is denoted by w_{i0t} .

Following Pissarides and Weber (1989), the log-linear Engel curve between food expenditure and income is estimated by:

$$\ln c_{it} = \alpha + X_{it} \Omega + \beta \ln y_{it}^P + \mu_t + \dot{Q}_{it} \quad (3)$$

where c_{it} , y_{it}^P , and X_{it} are food consumption, the permanent family income, and a vector of characteristics of household i at time t ; μ_t is a time effect; \dot{Q}_{it} represents the cumulative effects of other unobserved determinants of food consumption.

According to Hurst and Pugsley (2010), the gross permanent income y_{it}^P can be estimated from y_{it}^T by the following equation:

$$\ln y_{it}^T = \ln y_{it}^P + X_{it}\Gamma + \xi_{it} \quad (4)$$

Combining equations 2–4, we have:

$$\ln c_{it} = \alpha + X_{it}\Phi + \beta[\ln y_{it} + \ln(w_{i0t} + \sum_{k=1}^6 \delta_k w_{ikt})] + \mu_t + \zeta_{it} \quad (5)$$

where $\zeta_{ikt} = \hat{\alpha}_{ikt} - \beta\xi_{it}$, which can bias the estimated coefficient of β . To solve this problem, we follow Hurst and Pugsley's (2010) method by averaging household income and food expenditure both forward and backward in time. To estimate this equation, we use unweighted non-linear least squares regressions.⁶

3.2 Impact of income misreporting on inequality: a microsimulation model

From equation 5, we estimate all δ_k . Assuming that the same under-reporting rate is systematically applied to all households with similar income sources, the gross true household income of each family can be calculated based on equation 1. We use a static microsimulation SOUTHMOD model with an optimization procedure on top to estimate PIT and tax liability for both cases: reported vs true income. SOUTHMOD is based on the EUROMOD platform,⁷ which is well known as an EU-wide tax-benefit model. It applies user-defined tax and benefit policy rules to micro-data on individuals and households to estimate how much tax households need to pay, how many benefits they can receive, and finally their disposable income. The current version of SOUTHMOD for Viet Nam fundamentally focuses on tax aspects, including VAT, ad valorem tax, and PIT, and has not been updated for the current year policy reforms, which are primarily driven by tax rule changes for non-farm household businesses.

In this paper, while updating SOUTHMOD to reflect the current PIT policies, we apply the optimization procedures on top of this microsimulation model. This is because although there is no differentiation in the way wages or salaries are deducted and taxed, family bargaining or behaviour changes towards minimizing the family tax liabilities can occur through the decisions made by family members on how many dependents a taxpayer in dual-earner families should declare. To capture the optimization process, we need to recognize the associated families, who include at least one dependent person and two adults, with each earning at least VND9 million per month from wages or salaries, the tax deduction amount for any taxpayer. In these families, we first identify who is qualified as a dependant for tax purpose and then assign this dependent person to any eligible taxpayer who can be his/her child, spouse, or parent. Each taxpayer can claim one, two, or all qualified dependents. If a dependant is assigned to a taxpayer, other taxpayers in the same family cannot declare this dependent. A full combination of choices for each household is calculated, and the minimum value of family tax liability is selected as the ultimate option of this household. Assuming that tax evasion originates from income misreporting only, the discrepancy

⁶ We use the command *nl* in STATA.

⁷ See www.euromod.ac.uk for more details.

in tax liabilities calculated from the reported and the actual incomes indicates the scale of tax evasion.

Another limitation of the current SOUTHMOD is that cash transfer policy rules have not been modelled. However, this issue is mitigated by the fact that the eligibility requirements for many cash transfer policies are means-tested rather than income-tested (e.g. social benefits for families with revolutionary merits or those facing natural disasters). For other transfer policies, although the beneficiaries are mainly classified as poor, the correlation between income and poverty status is proved not to be strong. More than 50 per cent of the poor households identified by local authorities are not poor regarding income or expenditure measures (see Nguyen and Tran 2014). This is partly because of the method through which the poverty list was implemented, where poor households were identified in the first stage based on their ownership status of basic assets. Then income information was sought in the second stage only for the remaining households who were not classified as poor.⁸ In addition, rather than being solely based on the answers from households themselves, local authorities will decide who should be on the poor list because they tend to be better placed to accurately identify and target poor people and their problems (van de Walle 2004). Thus, income under-reporting is not expected to significantly affect the allowances households can receive and therefore their disposable income. After estimating tax liabilities and disposable incomes from both reported and actual incomes, we calculate tax revenue loss as a percentage share of actual tax revenue, and changes in Gini index and in median income ratios of the income distribution (P90/P10, P90/P50, P50/P10, P75/P25). The income is equivalized with the OECD-modified equivalence scale. This scale, first proposed by Hagenaars et al. (1994), assigns a value of 1 to the household head, 0.5 to each additional adult member, and 0.3 to each child.

4 Data and variables

This paper uses four rounds of the VHLSS in 2010, 2012, 2014, and 2016. These were conducted by the General Statistical Office of Viet Nam with technical support from the World Bank. They used the master sample drawn from the 15 per cent sample of the 2009 Housing and Population Census.⁹ In each survey, about 3,000 communes (700 from urban areas and 2,300 from rural areas) were randomly picked from the master sample.¹⁰ In each selected commune, one village and three households in that village were randomly chosen, creating a total number of observations of around 9,000 households for each survey. The VHLSS samples are therefore considered as stratified random samples and representative for the national, rural and urban, and regional levels.

It is worth noting that in each VHLSS, 50 per cent of communes are randomly selected for sample rotating purposes and all households in these selected communes were resampled in the succeeding VHLSS. Although the attrition rate in the panel data of the VHLSS is around 8 per

⁸ See details in Document No. 3385/LDTBXH-BTXH dated 30 September 2010, ‘Guideline on Poverty Census according to 1752/CT—TTg’

⁹ The 15 per cent sample was designed to be representative for all districts of Viet Nam (including 132 urban districts, 294 delta and coastal districts, and 256 mountain and island districts). Base weights were calculated as the inverse of the district sampling rates. See www.gso.gov.vn/Modules/Doc_Download.aspx?DocID=12608 for more information.

¹⁰ In the master sample, 5,490 communes (4,173 from rural areas and 1,317 from urban areas) were selected using the equal interval principle or every- k th rule (Hansen and Dang 2013) from the full list of Vietnamese communes, which was stratified on 63 provinces in both urban and rural areas (or 126 strata). These sampled communes were allocated over strata proportional to the square root of the number of households.

cent mainly due to migration (Nguyen 2016), households in the panel data are still representative (see Baulch and Vu 2010). Under this rotating method, around 3,800 households are sampled three times in four surveys, which allows us to follow Hurst and Pugsley's (2010) method of measuring permanent income and food expenditure by taking three-year averages of these variables in two sets of surveys (the first three surveys 2010, 2012, 2014 and the last three surveys 2012, 2014, 2016). Then equation 5 is estimated using these averages of income and expenditure together with household characteristics at time $t = 2012$ and $t = 2014$. The four rounds of VHLSS were generally collected in four quarters (March, June, October, December).¹¹ To control these time differences in all four years of the surveys, we apply the monthly consumer price index (CPI) to adjust all monetary values to December 2016 prices.

The VHLSSs classify a household as a person or group of people sharing accommodation and meals for at least 6 months over the 12 months prior to the time of interview. These people also share a common incoming and outgoing budget. Among these identified households, families with three generations (grandparents, parents, and children) and couples with independent children are very common in Viet Nam.

These surveys collect income information of all household members who are aged six years or older; working for other household businesses, state, private, or foreign firms; or self-employed in farming activities or non-farm production and business. Regarding wage and salaried jobs, the surveys obtain individuals' information on their main jobs (which occupies most of their working time) for the last 12 months, their second wage job, and other jobs. Income from these jobs covers not only wages and salaries paid in the previous 12 months, but also a total amount of work-related allowances such as bonuses, allowances for uniforms, lunch, business trips, sickness, labour accidents, pregnancy, and so on. Although some income items containing payments for uniforms, telephones, and stationery are tax-exempt and subjected to caps, we cannot differentiate them. Therefore, all of these allowances are considered as taxable income when calculating the PIT. The survey also gives information on types of employers: whether they are other household business, SOEs, private firms, or FIEs. There are two kinds of household businesses taken into account: one with business registration and the other without this licence.

Regarding self-employment income, in addition to the availability of output classification by industry for tax calculation purposes, the surveys also gather income information on gross output produced and total product sold or exchanged in both farm and non-farm activities within the last 12 months. In this paper, we use income from gross production output rather than profits from sales or exchanges of other goods because the former extends to cover the output that is self-supplied and consumed by households. This definition helps to mitigate an issue raised by previous studies that farmers should behave differently in terms of food expenditure compared to others when the output values of self-supply and self-consumption are excluded in their calculations of income and food expenditure.

Besides wages, salaries and self-employment income, the surveys also collect other taxable income data, such as rental income (from leasing land, housing, or production assets); investment income (interest on deposits, stocks, shares, or lending). Additionally, they gather information on non-taxable income, including remittances (gifts or cash or in-kind assets or non-asset goods from overseas or within the country), pensions, allowances, or social benefits (unemployment allowance, severance pay, allowance for working capacity loss, families with revolutionary merits or

¹¹ Due to time constraints, the 2010 VHLSS was carried out in June, October, and December of 2010.

beneficiaries of social policies, or assistance to overcome natural disasters and fire). Insurance compensation and donations are also incorporated in the surveys.

Using this very rich information on income, the accurately reported part of income y_{it} in equation 5 consists of income from farming output produced, remittances (excluding gifts of assets), pensions, other non-taxable income (allowances, social benefits, insurance compensation, donations), and wages from FIEs. The misreported part of income, therefore, includes wages and salaries (with work-related allowances) received from working for other household businesses, state and private firms, income from unregistered and registered non-farm businesses, rental income, and investment income.

In addition to income information, the VHLSSs also collect food and non-food expenditure. In this paper, food consumption is suitable for measuring the Engel curve for a number of reasons. First, individuals tend not to be afraid of reporting food expenditure correctly. Second, because food is often bought, food expenditure is most likely better approximated by 30-day purchases than other expenditures. Third, food is less likely to be registered as a business expense, reducing the probability of consumption measurement errors. Information on food expenditure includes both purchased, exchanged, or self-made products catering to the household's needs on festive occasions and every day. For eating and drinking outside the home, only the meals that the household has to pay for are counted.

Finally, the VHLSSs contain detailed information on household and individual characteristics such as demography (age, sex, ethnicity, marital status), employment and labour force participation, education, and health. The list of dependent and independent variables to estimate equation 5 is summarized in Appendix A. We then use the sampling weights to calculate tax evasion, Gini index, and income ratios at the national level.

5 Results

In this section we first examine the distribution of income components and food expenditure (Table 1), as well as household characteristics and spatial diversity (Table 2) between cross-sectional and panel data to check for their comparability. We then analyse this distribution among income quintile groups that are computed from the total equivalized pre-tax income attributed to each member of the household (Tables 3 and 4). This analysis helps us to understand the income distribution in Viet Nam. Next, we report the outcome from unweighted non-linear least square estimations of income under-reporting rates (Table 5) together with the compliance rates (Table 6) for different income sources. We also have a quick look at income and expenditure redistribution among income quintile groups when true income is applied (Table 7). Finally, we estimate the impact of income under-reporting on the size of tax evasion (Table 8) and the change in inequality (Table 9) in Viet Nam.

Table 1 decomposes income and food expenditure by sources in single years of 2012 (column 2) and 2014 (column 4) as well as three-year average panel data (columns 3 and 5) collected from the VHLSS. There were a total of 9,384 households surveyed in 2012 and 9,377 households in 2014. Of these, one balanced panel of 1,948 households was interviewed repeatedly in all three years of 2010, 2012, and 2014, and another balanced panel of 1,895 households was interviewed in 2012, 2014, and 2016. These two balanced panels will be pooled together to create a panel of 3,843 observations that are applied to estimate equation 5 (Table 5).

Table 1: Household income and food expenditure by sources: cross-sectional vs panel data

	2012	2010–14 panel	2014	2012–16 panel
Expenditure on food, beverages, and eating out	3,616 [2,423; 41.6]	4,041 [2,427; 46.4]	5,137 [3,861; 54.6]	4,696 [3,062; 49.2]
Expenditure on food	2,568 [1,400; 29.5]	2,800 [1,287; 32.1]	3,359 [1,822; 35.7]	3,073 [1,454; 32.2]
Expenditure on beverages and eating out	1,048 [1,540; 12.0]	1,241 [1,584; 14.2]	1,778 [2,868; 18.9]	1,623 [2,149; 17.0]
Household income	8,701 [10,616; 100.0]	8,713 [7,078; 100.0]	9,403 [9,629; 100.0]	9,540 [8,416; 100.0]
Wages in household businesses	1,071 [2,080; 12.3]	1,054 [1,622; 12.1]	1,112 [2,167; 11.8]	1,181 [1,923; 12.4]
Wages in SOEs	1,192 [3,697; 13.7]	1,144 [2,985; 13.1]	1,379 [3,968; 14.7]	1,339 [3,616; 14.0]
Wages in private firms	910 [3,159; 10.5]	818 [2,210; 9.4]	965 [3,027; 10.3]	971 [2,541; 10.2]
Wages in FIEs	297 [1,864; 3.4]	264 [1,226; 3.0]	407 [2,117; 4.3]	383 [1,502; 4.0]
Income from farming activities	2,021 [4,354; 23.2]	2,395 [4,108; 27.5]	2,087 [4,249; 22.2]	1,994 [3,482; 20.9]
Income from formal non-farm prod	955 [4,318; 11.0]	967 [4,179; 11.1]	1,165 [6,139; 12.4]	1,251 [4,273; 13.1]
Income from informal non-farm prod	941 [2,969; 10.8]	935 [2,307; 10.7]	952 [2,967; 10.1]	955 [2,235; 10.0]
Rental and investment income	350 [3,375; 4.0]	227 [800; 2.6]	315 [2,015; 3.3]	360 [2,786; 3.8]
Pension and social benefits	379 [2,579; 4.4]	353 [970; 4.0]	417 [1,591; 4.4]	437 [1,291; 4.6]
Remittances	552 [1,606; 6.3]	524 [1,042; 6.0]	563 [1,686; 6.0]	629 [1,415; 6.6]
Observations	9,384	1,948	9,377	1,895

Notes: (1) Numbers in brackets report standard deviations (on the left) and the percentage share of household income by sources (on the right); (2) values in thousands of constant December 2016 VND; (3) columns 2 and 4 provide the mean values of monthly household income and expenditure; (4) columns 3 and 5 measure three-year averages of monthly income and expenditure.

Source: Authors' calculation based on VHLSSs 2010, 2012, 2014, and 2016.

In general, as shown in Table 1, there is not a dramatic difference between cross-sectional and panel samples. The discrepancy between these samples and the smaller standard deviations in the panel sample compared to cross-sectional data refer to the existence of transitory variation (columns 2 and 4) around permanent income (columns 3 and 5), in addition to measurement error in reported income. Using the permanent income and expenditure variables from the panel data, therefore, mitigates the problem of underestimating the expenditure–income elasticity β (see Hurst and Pugsley 2010).

On average, around 45–50 per cent of household income is spent on food, beverages, and eating out, of which about two-thirds represents food only. For simplicity, we use total expenditure on food, beverages, and eating out as our primary dependent variable to estimate the income under-reporting rates of equation 5. We use this variable because food is considered as a relative necessity compared to other expenditure categories in many empirical studies. When more aggregate consumption measures are used, income under-reporting rates can be under-estimated (see Hurst and Pugsley 2010).

Regarding income components in Table 1, on average, wage income accounts for approximately 40 per cent of pre-tax household income¹² reported in the surveys. Of this, wages earned from working for household businesses (HBs), public (SOEs), and private sectors share a relatively equal proportion of 10–14 per cent of total income. Wages from foreign sectors (FIEs) only explain 3–4 per cent of total income. Income share from agricultural activities in Viet Nam is sizeable, with 20–30 per cent of total income on average, which makes Viet Nam different from developed countries. This fact explains why we cannot drop households with farming income from our Vietnamese study. Non-farm income also plays a moderate role, accounting for 22–23 per cent of total income. Another 3–4 per cent of income is from rental and investment activities and the rest (10–11 per cent) is from remittances, pensions, and social transfers. From this income distribution, we can see clearly the multiple-source-based income characteristics in Viet Nam, suggesting the application of non-linear least squares estimation (applied by Feldman and Slemrod (2007) and Hurst and Pugsley (2010)) rather than the simple linear least squares regression of the Engel curve pioneered by Pissarides and Weber (1989). Considering tax evasion as the only reason to under-report income, we group tax-free income sources including farm income, remittances, pensions, and other social benefits together with wage income from foreign sectors as the portion of income accurately reported. Wages earned in foreign companies tend to be correctly reported because foreign firms in Viet Nam often have to comply with more stringent accounting principles and are rarely involved in collusion with tax authorities.¹³

A cross-check between cross-sectional and panel data is also implemented regarding household characteristics and spatial distributions (Table 2). Compatible values are shown between these two samples. From Table 2, we observe that the age of a household head is about 50 on average and a dominant proportion of the population (83 per cent) are Vietnamese or Chinese. About 55 per cent of household heads work full-time (at least 30 hours per week), 20 per cent work part-time (15–30 hours per week), and 10 per cent work on a casual basis (at least one hour during the year). This leaves 14 per cent of household heads not working. The average number of working members (at least 15 hours per week) is two, and the average size of a family is approximately four members. This is because in Viet Nam three-generation families where the head or spouse take care of their

¹² Pre-tax household income includes market income before tax and social transfers.

¹³ We also run a regression separating wages in FOEs as an income source under-reported. However, the estimated compliance rate of this income type is very close to 100 per cent, which rejects the assumption of misreporting happening to this source.

parents or their grandchildren are prevalent, and an employee can get the tax deduction on not only their dependent child or spouse but also their parents, which is different from the situation in developed countries.

Table 2: Household characteristics and regional distribution: cross-sectional vs panel data

	2012	2010–14 panel	2014	2012–16 panel
Age of household (HH) head	49.75 (14.24)	49.96 (14.11)	50.74 (14.13)	51.32 (13.98)
HH head is Vietnamese or Chinese	0.83 (0.38)	0.82 (0.39)	0.83 (0.37)	0.83 (0.37)
HH head is unemployed	0.14 (0.35)	0.14 (0.35)	0.15 (0.35)	0.15 (0.36)
HH head works full-time	0.55 (0.50)	0.57 (0.50)	0.56 (0.50)	0.55 (0.50)
HH head works part-time	0.21 (0.40)	0.20 (0.40)	0.20 (0.40)	0.20 (0.40)
HH head works on a casual basis	0.10 (0.30)	0.10 (0.29)	0.09 (0.29)	0.09 (0.29)
Number of children aged under 2	0.11 (0.33)	0.11 (0.33)	0.12 (0.34)	0.10 (0.32)
Number of children aged 2–5	0.25 (0.49)	0.25 (0.48)	0.24 (0.49)	0.23 (0.47)
Number of children aged 6–10	0.31 (0.56)	0.31 (0.56)	0.32 (0.55)	0.31 (0.54)
Number of children aged 11–15	0.33 (0.58)	0.34 (0.60)	0.30 (0.56)	0.31 (0.55)
Number of adults aged 16+	2.90 (1.22)	2.94 (1.25)	2.85 (1.19)	2.84 (1.17)
Number of working members	2.01 (1.13)	2.07 (1.17)	1.99 (1.11)	1.98 (1.13)
Red River delta	0.21 (0.41)	0.21 (0.41)	0.21 (0.41)	0.22 (0.41)
Mountainous areas	0.18 (0.38)	0.18 (0.38)	0.18 (0.38)	0.19 (0.40)
Coastal central region	0.22 (0.41)	0.23 (0.42)	0.22 (0.41)	0.22 (0.41)
Central highlands	0.07 (0.25)	0.07 (0.26)	0.07 (0.25)	0.07 (0.26)
South-east	0.12 (0.32)	0.11 (0.31)	0.12 (0.32)	0.11 (0.31)
Mekong River delta	0.20 (0.40)	0.20 (0.40)	0.20 (0.40)	0.19 (0.39)
Urban areas	0.29 (0.45)	0.27 (0.45)	0.30 (0.46)	0.31 (0.46)
Observations	9,384	1,948	9,377	1,895

Notes: (1) Standard deviations are reported in brackets; (2) columns 2 and 4 shows statistics of all households in the years 2012 and 2014; (3) columns 3 and 5 report statistics of the panel groups in the years 2012 and 2014.

Source: Authors' calculation based on VHLSSs 2010, 2012, 2014, and 2016.

Regarding spatial distribution, about 70 per cent of the population lives in rural areas, explaining why a moderate portion of the income comes from farming activities. Our sample is representative at provincial and regional levels. Here, 40 per cent reside in two delta regions, 25 per cent in

mountainous areas and central highlands, more than 20 per cent in the coastal central regions and the rest (above 10 per cent) living in south-east, the top income region in Viet Nam (see Table 2).

Tables 3 and 4 compare per capita income components, expenditure, and household characteristics across income quintile groups in 2016. Sampling weights are applied to obtain nationally representative statistics. The income quintile groups are computed on the basis of the total equivalized pre-tax income attributed to each member of the household, of which household size is adjusted using the OECD-modified equivalence scale.

Table 3: Per capita income and food expenditure by income quintiles, 2016

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Expenditure on food, beverages, and eating out	1,568	2,123	2,614	3,232	4,730
	(99.11)	(68.96)	(57.15)	(49.89)	(38.29)
Expenditure on food	1,251	1,526	1,716	1,959	2,493
	(79.02)	(49.55)	(37.53)	(30.23)	(20.18)
Expenditure on beverage and eating out	318	598	898	1,273	2,237
	(20.08)	(19.41)	(19.63)	(19.65)	(18.11)
Per capita income	1,583	3,079	4,573	6,478	12,353
	(100)	(100)	(100)	(100)	(100)
Wages in HBs	352	746	955	867	566
	(22.24)	(24.24)	(20.88)	(13.39)	(4.58)
Wages in SOEs	27	184	403	917	1,990
	(1.71)	(5.99)	(8.81)	(14.15)	(16.11)
Wages in private firms	72	316	588	1,086	1,861
	(4.57)	(10.25)	(12.86)	(16.76)	(15.06)
Wages in FIEs	13	68	278	575	674
	(0.84)	(2.21)	(6.07)	(8.87)	(5.45)
Income from farming activities	686	859	914	919	1,849
	(43.35)	(27.88)	(19.99)	(14.19)	(14.97)
Income from formal non-farm production	33	153	326	666	2,481
	(2.06)	(4.97)	(7.12)	(10.28)	(20.09)
Income from informal non-farm production	69	229	552	809	1,427
	(4.33)	(7.45)	(12.07)	(12.48)	(11.55)
Rental and investment income	30	62	89	136	571
	(1.87)	(2.00)	(1.95)	(2.11)	(4.62)
Pension and social benefit	104	153	146	162	254
	(6.60)	(4.98)	(3.19)	(2.51)	(2.06)
Remittances	178	280	300	326	662
	(11.26)	(9.11)	(6.57)	(5.04)	(5.36)

Note: (1) Numbers in brackets report the percentage share of pre-tax income by sources; (2) values in thousands of constant December 2016 VND; (3) sampling weights are applied to obtain nationally representative statistics; (4) income quintile groups are computed based on total equivalized pre-tax income attributed to each member of the household, of which household size is adjusted using the OECD-modified equivalence scale.

Source: Authors' calculation based on VHLSS 2016.

In Table 3 we find the disparities in components of reported income among income groups. Poorer households tend to rely mainly on tax-free income sources including farming income, remittances, pensions, and social benefits (61 per cent of pre-tax household income for the bottom quintile group), while richer households tend to earn more from wages or salaries in the public

and private sectors, and non-farm production income (63 per cent of income for the top quintile group). Additionally, while a higher share of wage income in the household business sector is observed among the poorer households, the higher-income groups tend to rely more on rental and investment income as well as earnings from the foreign sector. Regarding expenditure on food, beverages, and eating out, the poorest spend their reported income entirely on food expenditure, which suggests an income under-reporting issue even within this group. Besides that, the food expenditure share of income reduces when income is higher, implying their Engel curve relationship. It is further supported by strong evidence in Figure 2 showing that the outcome from the non-parametric model demonstrates a linear correlation between income and expenditure under the logarithm forms.

Figure 2: Non-parametric estimate of food Engel curve using three-year VHLSS panel



Source: Authors' calculations based on VHLSS 2012, 2014, and 2016.

These income groups are also associated with different characteristics. As described in Table 4, heads of wealthier households tend to be younger, belong to ethnic majority groups (including Vietnamese and Chinese), and are more likely to work in full-time jobs. These more prosperous families also have fewer young dependent children, more working members, and tend to live in high-income regions including the south-east and Red River delta, and in urban areas. These characteristics, combined with the income distribution of Table 3, show that the old and pensioner household heads tend to live with their grandchildren in more impoverished families and are more likely to receive a pension, other social benefits, and private transfers.

Table 4: Household characteristics and regional distribution by income quintiles, 2016

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Age of household (HH) head	52.01 (15.66)	52.41 (13.52)	51.31 (12.74)	50.17 (12.49)	49.57 (12.45)
HH head is Vietnamese or Chinese	0.53 (0.50)	0.83 (0.38)	0.91 (0.28)	0.96 (0.20)	0.96 (0.19)
HH head is unemployed	0.15 (0.36)	0.15 (0.36)	0.16 (0.37)	0.15 (0.35)	0.16 (0.37)
HH head works full-time	0.38 (0.49)	0.48 (0.50)	0.60 (0.49)	0.66 (0.47)	0.70 (0.46)
HH head works part-time	0.32 (0.47)	0.25 (0.43)	0.18 (0.38)	0.13 (0.34)	0.10 (0.30)
HH head works on a casual basis	0.15 (0.36)	0.12 (0.32)	0.06 (0.25)	0.06 (0.24)	0.04 (0.19)
Number of children aged under 2	0.17 (0.39)	0.17 (0.40)	0.16 (0.38)	0.13 (0.35)	0.12 (0.33)
Number of children aged 2–5	0.40 (0.63)	0.38 (0.59)	0.31 (0.52)	0.33 (0.56)	0.32 (0.55)
Number of children aged 6–10	0.35 (0.62)	0.33 (0.56)	0.35 (0.57)	0.39 (0.59)	0.48 (0.64)
Number of children aged 11–15	0.37 (0.64)	0.35 (0.59)	0.37 (0.61)	0.36 (0.60)	0.40 (0.61)
Number of adults aged 16+	3.08 (1.31)	3.22 (1.19)	3.26 (1.14)	3.24 (1.20)	3.20 (1.24)
Number of working members	2.00 (1.37)	2.15 (1.12)	2.30 (1.03)	2.39 (1.01)	2.40 (1.05)
Red River delta	0.12 (0.33)	0.20 (0.40)	0.25 (0.43)	0.28 (0.45)	0.32 (0.46)
Mountainous areas	0.30 (0.46)	0.14 (0.35)	0.11 (0.31)	0.09 (0.29)	0.06 (0.24)
Coastal central region	0.30 (0.46)	0.28 (0.45)	0.23 (0.42)	0.16 (0.37)	0.13 (0.33)
Central highlands	0.09 (0.29)	0.08 (0.27)	0.06 (0.23)	0.05 (0.22)	0.05 (0.22)
South-east	0.02 (0.14)	0.07 (0.26)	0.15 (0.36)	0.25 (0.43)	0.30 (0.46)
Mekong River delta	0.17 (0.37)	0.22 (0.42)	0.20 (0.40)	0.16 (0.37)	0.14 (0.34)
Urban areas	0.09 (0.28)	0.19 (0.40)	0.30 (0.46)	0.42 (0.49)	0.59 (0.49)

Notes: (1) Standard deviations are reported in brackets; (2) sampling weights are applied to obtain nationally representative statistics; (3) income quintile groups are computed based on total equivalized pre-tax income attributed to each member of the household, of which household size is adjusted using the OECD-modified equivalence scale.

Source: Authors' calculations based on VHLSS 2016.

Now we move to the estimation of equation 5 using the non-linear least squares regression. Tables 5 and 6 report the direct outcome and a calculation of the compliance rates of this estimation using the three-year average panel data for 2012 and 2014, which were described in Tables 1 and 2. Table 5 presents the coefficient parameters of income under-reporting rates (δ_k), the log of income (β) and household characteristics in column 2, and their 95 per cent confidence intervals in column 3. Recall that if $\delta_k = 1$, we do not expect an under-reporting issue for income k . When the assumption

that $\delta_k > 1$ is not rejected, we can conclude that the income k tends to be under-reported by a rate of $(\hat{\delta}_k - 1) / \hat{\delta}_k \times 100$ (per cent). These rates are reported in Table 6.

The results in Table 5 are striking. The estimated factors $\delta_k > 1$ are well above the value of 1 in terms of the 95 per cent confidence intervals—that is, all lower bounds of these estimates are greater than 1. After controlling for household characteristics, spatial, and time differences, we find that not only self-employment income and capital income but also wage income from public and private sectors are under-reported, which is different from the assumption made by previous studies on developed countries.

Table 5: Non-linear least squares estimation of income under-reporting

	Coefficients	95 per cent confidence interval
$\hat{\beta}$, ln(Total income)	0.483***	[0.465, 0.502]
$\hat{\delta}_1$, wages in HBs (ratio)	1.116***	[1.004, 1.228]
$\hat{\delta}_2$, wages in SOEs (ratio)	1.661***	[1.476, 1.846]
$\hat{\delta}_3$, wages in private firms (ratio)	1.489***	[1.301, 1.677]
$\hat{\delta}_4$, income from formal non-farm production (ratio)	1.208***	[1.057, 1.359]
$\hat{\delta}_5$, income from informal non-farm production (ratio)	1.300***	[1.146, 1.454]
$\hat{\delta}_6$, rental and investment income (ratio)	1.610***	[1.189, 2.031]
Age of household (HH) head	-0.002***	[-0.003, -0.001]
HH head is Vietnamese or Chinese	0.030*	[-0.005, 0.065]
HH head is unemployed	0.079***	[0.037, 0.121]
HH head works full-time	0.080***	[0.039, 0.121]
HH head works part-time	0.087***	[0.044, 0.130]
Number of children aged under 2	0.030*	[-0.001, 0.061]
Number of children aged 2–5	0.062***	[0.041, 0.083]
Number of children aged 6–10	0.072***	[0.053, 0.090]
Number of children aged 11–15	0.085***	[0.068, 0.102]
Number of adults aged 16+	0.119***	[0.106, 0.131]
Number of working members	-0.039***	[-0.053, -0.024]
Mountainous areas	0.039**	[0.003, 0.076]
Coastal central region	0.035**	[0.005, 0.065]
Central highlands	-0.062***	[-0.105, -0.018]
South-east	-0.028	[-0.064, 0.009]
Mekong River delta	-0.054***	[-0.085, -0.024]
Urban areas	0.131***	[0.106, 0.155]
Year = 2014	0.106***	[0.087, 0.126]
Constant	3.441***	[3.277, 3.604]
Observations	3,843	
Adjusted R^2	0.728	

Notes: (1) * $p \leq 0.10$; ** $p \leq 0.05$; *** $p \leq 0.01$; (2) column 2 reports lower and upper bounds of 95 per cent confidence intervals; (3) results from this table are non-linear least squares estimates of equation 5.

Source: Authors' calculation based on three-year VHLSS panels.

Notably, *ceteris paribus*, the estimated parameter for income from working for SOEs becomes largest (1.66). This means that the wage income from public sector income reported in the surveys needs to be scaled up by a factor of 1.66 to correct for under-reporting of this income source. This outcome is further supported by the narrow band of the 95 per cent confidence interval around 1.66. This number corresponds to an average under-reporting rate of 60 per cent of the true income (see Table 6).

Table 6: Under-reporting rates of the true income by sources

	Mean	Lower bound	Upper bound
$\hat{\delta}_1$, wages in HBs (ratio)	10.41	0.41	18.59
$\hat{\delta}_2$, wages in SOEs (ratio)	39.79	32.25	45.83
$\hat{\delta}_3$, wages in private firms (ratio)	32.83	23.13	40.37
$\hat{\delta}_4$, income from formal non-farm prod (ratio)	17.2	5.37	26.41
$\hat{\delta}_5$, income from informal non-farm prod (ratio)	23.08	12.77	31.2
$\hat{\delta}_6$, rental and investment income (ratio)	37.89	15.92	50.75

Notes: (1) The percentage share of income under-reporting is $(\hat{\delta}_k - 1) / \hat{\delta}_k \times 100$; (2) columns 2 and 3 report lower and upper bounds of these shares.

Source: Authors' calculation based on three-year VHLSS panels.

Also demonstrated in Table 5, ranking second place for non-compliance rate is wages from the private sector, with a factor of 1.49 and a 95 per cent confidence interval between 1.30 and 1.68, corresponding to the average non-compliance rate of 33 per cent in Table 6. Although the mean value of the estimated factor for capital income is larger (1.61), the band of the 95 per cent confidence interval is relatively large, from 1.19 to 2.03 (equivalent to under-reporting rates between 16 and 51 per cent in Table 6). For self-employment income, the estimated factors are 1.3 for unregistered businesses and 1.21 for registered activities (corresponding to the average non-compliance rates of 17 per cent and 23 per cent in Table 6). We also see a smaller income under-reporting rate (10 per cent at mean values) from wages in HBs.

This popularity of income under-reporting and tax evasion in many sectors in Viet Nam may be the result of prevailing corruption, one of the most critical issues in doing business facing Viet Nam at present. Taxation is perceived as the sixth most corrupt sector. According to the Viet Nam Chamber of Commerce and Industry (VCCI), among 137 countries, Viet Nam ranked 81st in corruption and morality, and 109th in bribery and informal payments. Giang et al. (2016), in their survey of 525 HBs in all regions of Viet Nam, find that about 70 per cent of the respondents indicated that they always or often colluded with tax inspectors for mutual benefit. According to Nguyen et al. (2017), tax corruption may result from political factors (lack of transparency or accountability), institutional factors (lack of a competent public sector), cultural factors (sharing within the extended family of the local community, a compromising and paternalistic culture), and economic factors (low wages in the public sector where public servants cannot live on their official salaries). In addition to the complex and discrete tax law, there is no effective system of independent tax dispute resolutions, poor accounting and record-keeping, and strong business motives to pay as little tax as possible, which can also result in widespread tax corruption (Nguyen et al. 2017).

Overall, the outcome shown in Tables 5 and 6 suggests substantial non-compliance exists. We reject the null hypothesis that the coefficients equal 1, meaning full compliance. Thus, the larger portion of a household's income that comes from earnings in public or private sectors, self-employment income from non-farm activities, or capital income, the larger is the share of income that is concealed. Using these estimates, and assuming that tax non-compliance occurs in these sectors only, it is possible to gauge overall hidden income in the Vietnamese economy.

Table 7 provides the compliance rates of total income as well as the distribution of true income across income quintile groups. This table is equivalent to Table 3 but applies to the estimated actual income. Sampling weights are used to obtain nationally representative statistics. Results of compliance rates (reported income share of correct income) from Table 7 show that wealthier households tend to under-report their income at a higher rate: 23 per cent for the top quintile, 17 per cent for the middle quintile, and 7 per cent for the lowest part of the income distribution. Again, it is because the poorest rely mainly on tax-free income sources while the income of the richest comes largely from wages and non-farm activities. Thus, on average, the officially reported income is 20 per cent lower than the true one. This actual income means the correlation between food expenditure and income reduces by about 10 percentage points compared to the reported income (see Tables 3 and 7).

Table 7: Estimated true income by sources and income quintiles, 2016

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5
Reported food expenditure per capita	1,568 (92.08)	2,123 (59.78)	2,614 (47.49)	3,232 (39.58)	4,730 (29.72)
Reported income per capita	1,583 (92.91)	3,079 (86.68)	4,573 (83.10)	6,478 (79.34)	12,353 (77.60)
Estimated true income per capita	1,703 (100)	3,552 (100)	5,503 (100)	8,165 (100)	15,918 (100)
Farm income, pensions, remittances, allowances and wages in FIEs	982 (57.64)	1,360 (38.30)	1,638 (29.77)	1,983 (24.28)	3,439 (21.60)
Wages in HBs	393 (23.06)	833 (23.45)	1,066 (19.37)	968 (11.85)	632 (3.97)
Wages in SOEs	45 (2.64)	306 (8.62)	669 (12.17)	1,523 (18.65)	3,306 (20.77)
Wages in private firms	108 (6.32)	470 (13.23)	876 (15.91)	1,617 (19.80)	2,770 (17.40)
Income from formal non-farm production	39 (2.32)	185 (5.21)	393 (7.14)	804 (9.85)	2,997 (18.83)
Income from informal non-farm production	89 (5.23)	298 (8.39)	717 (13.04)	1,051 (12.87)	1,855 (11.65)
Rental and investment income	48 (2.79)	99 (2.79)	143 (2.60)	220 (2.69)	919 (5.78)

Note: (1) Numbers in brackets report the percentage share of pre-tax income by sources; (2) values in thousands of constant December 2016 VND; (3) sampling weights are applied to obtain nationally representative statistics; (4) income quintile groups are computed based on total equivalized pre-tax income attributed to each member of the household, of which household size is adjusted using the OECD-modified equivalence scale; (5) this table is similar to Table 3 but applied to estimated true income of which income component k is multiplied to $\hat{\delta}_k$.

Source: Authors' calculations based on VHLSS 2016.

In this paper, we take advantage of the SOUTHMOD microsimulation model to impute the scale of tax evasion (Table 8) and its distributive effects (Table 9). We also report tax evasion by income

components (Table 8), which are calculated based on the shares of each income component. We find that the total tax evasion originating from income under-reporting is almost VND23,000 billion (in December 2016 VND), with a 95 per cent confidence interval ranging from VND11,000 to 36,000 billion. Thus, the government could increase its PIT revenue by between 17 and 57 per cent, at a mean rate of 36 per cent.¹⁴ When decomposing this total tax evasion by income source, we observe that it is mainly from wage income in public and private sectors (62 per cent), followed by non-farm self-employment income (25 per cent) and capital income (18 per cent) (see Table 8).

Table 8: Tax evasion by income sources, 2016

	Mean	Lower bound	Upper bound
Total tax evasion	22,741	11,030	36,055
From wages in HBs	177	66	343
From wages in SOEs	7,994	4,839	11,862
From wages in private firms	6,132	3,139	9,882
From formal non-farm business	2,946	822	5,071
From informal non-farm business	2,822	1,342	4,373
From rental and investment income	2,670	822	4,524

Notes: (1) Values in billions of constant December 2016 VND; (2) sampling weights are applied to obtain nationally representative statistics; (3) columns 2 and 3 report lower and upper bounds of tax evasion; (4) tax evasion by income components is calculated based on the shares of each income component.

Source: Authors' calculations based on VHLSS 2016.

Table 9 reports the estimates of inequality indices from gross income or market income (i.e. income before tax and benefit), pre-tax income (i.e. market income and social benefit), and disposable income (i.e. after-tax income). These indices are computed on the basis of income reported in the 2016 VHLSS survey (column 1) and true income estimated from equation 5 at mean values (column 2) and their 95 per cent confidence intervals (lower bound in column 3 and upper bound in column 4). We calculate not only the Gini index but also the crucial income ratios of the income distribution (P90/P10, P90/P50, P50/P10, P75/P25).

Without tax evasion, on average, the Gini coefficient for disposable income would be 0.409, an increase of 0.03 from the one using reported income (0.379).¹⁵ This increase in the Gini index lies in a range between 0.021 and 0.038 when applying the 95 per cent confidence intervals. Similar patterns of Gini coefficients are shown among both market income and pre-tax income, supporting the conclusion that income misreporting tends to underestimate the true income inequality.

Intuitively, regarding the full income distribution, the Lorenz curve of true income at mean values, as demonstrated in Figure 3, is always lower than the one using reported income. This outcome is further supported by evidence of higher income ratios among correct values at not only mean values but also lower and upper bounds of their 95 per cent confidence interval in comparison with reported income. Thus, generally, tax evasion makes the distribution of reported income appear more equal than it is.

¹⁴ The total PIT revenue collected by the Vietnamese government in 2016 was approximately VND63,600 billion.

¹⁵ According to Benjamin et al. (2016), the estimates of the Gini index for reported pre-tax income per capita were 0.375 in 2012 and 0.360 in 2014. Our calculation of this index is higher, at 0.383 (for pre-tax income), because of two possible reasons. First, we use the income values of constant December 2016 VND instead of January 2012 VND. Second, we use the household size adjusted by the OECD-modified equivalence scale to calculate income per capita.

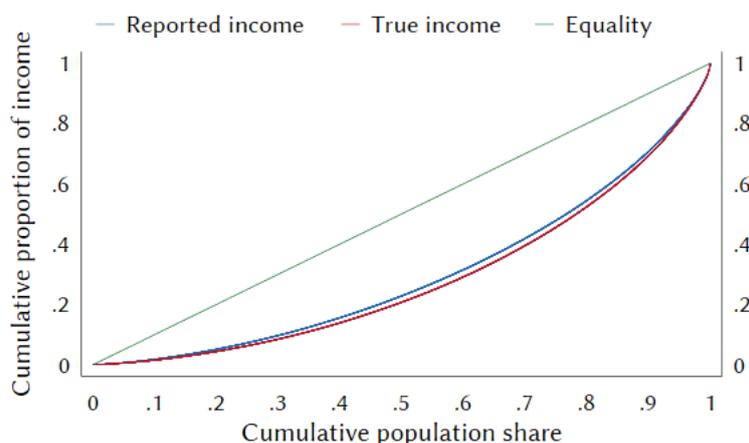
Table 9: Impact of income under-reporting and tax evasion on income distribution

	Reported	Mean	Lower bound	Upper bound
Gross income				
Gini	0.389	0.419	0.410	0.428
P90/P10	6.593	8.165	7.660	8.630
P90/P50	2.268	2.494	2.426	2.554
P50/P10	2.907	3.274	3.157	3.379
P75/P25	2.601	3.001	2.901	3.095
Pre-tax income				
Gini	0.383	0.414	0.404	0.423
P90/P10	6.257	7.663	7.205	8.180
P90/P50	2.243	2.470	2.405	2.536
P50/P10	2.789	3.102	2.996	3.226
P75/P25	2.537	2.923	2.825	3.035
Disposable income				
Gini	0.379	0.409	0.400	0.417
P90/P10	6.206	7.580	7.164	8.079
P90/P50	2.231	2.455	2.400	2.515
P50/P10	2.781	3.087	2.985	3.213
P75/P25	2.528	2.908	2.811	3.020

Notes: (1) Indices from column 1 are estimated based on the reported income (i.e. with tax evasion); (2) indices from columns 2–4 are estimated based on mean value and 95 per cent confidence interval of estimated true income from equation 5; (3) disposable income = gross income – tax + social benefit, which includes pre-95 pension and other benefits; (4) pre-tax income = gross income + social benefit.

Source: Authors' calculations based on VHLSS 2016.

Figure 3: Lorenz curve comparison: reported vs true disposable income in 2016



Source: Authors' calculations based on VHLSS 2016.

It is also worth noting that the role of the PIT and cash transfers in Viet Nam is minimal compared to developed countries. For example, at both reported and true values of income, the Gini indices reduce by only 0.01 when the PIT and cash transfer policies are taken into account (see Table 9). This reduction is, however, far from the outcome achieved in developed countries, such as Australia with a decline of 0.133 in Gini coefficient in 2016 (see Li and La 2018).

6 Conclusion

Essentially all empirical work using survey data assumes that household income is not systematically misreported. However, there is reason to believe that this assumption may not hold, especially in a country with widespread corruption like Viet Nam. Previous studies on this topic mainly focus on developed countries and assume that wage income is correctly reported and self-employed workers substantially under-report their income. However, what was less known is whether wage income is also under-reported and the extent to which each type of income is under-reported for the case of Viet Nam, a developing country relying on farming activities. Instead of assuming that salary workers honestly report all their incomes, we have also admitted the chance of hiding earnings by employees.

We apply the non-linear least squares method to estimate different compliance rates for different sources of income. Our method follows the approach suggested by Feldman and Slemrod (2007) and Hurst and Pugsley (2010). This method overcomes the approach used in many studies in which a range of cut-off levels of self-employment income is applied (25 per cent in Pissarides and Weber 1989; 30 per cent in Schuetze 2002; 0 per cent in Kukk and Staehr 2017; 20 per cent in Staehr 2004). In other words, we have proposed here a broader interpretation of the standard model of Pissarides and Weber (1989).

Using data from VHLSSs 2010, 2012, 2014, and 2016, we find that, on average, wage earnings from the public sector are under-reported at the highest rate (40 per cent of true income), followed by wages from private firms and capital income (33–38 per cent). Under-reporting is also found among self-employment income but at a lower rate (23 per cent and 17 per cent for unregistered and registered businesses, respectively). Wage income from other HBs are misreported at the lowest level (10 per cent). We also observe that this income under-reporting rate is highest among the richest (22 per cent) and lowest among the poorest (7 per cent). In general, the officially reported income only accounts for 80 per cent of the true income, leaving 20 per cent of true income misreported.

Consequently, our results are economically significant, indicating that tax liabilities would increase by about VND23,000 billion, accounting for 36 per cent of the PIT budget revenue in 2016. Of this, tax loss from wage income explains 62 per cent, followed by non-farm self-employment income (25 per cent) and capital income (12 per cent). The income under-reporting also implies an underestimation of the inequality measures when the reported income is used. The Gini coefficient for disposable income would increase from 0.379 (with income under-reporting) to 0.409 (without hidden income) and tax-benefit policies only play a minimal role in mitigating income inequality in Viet Nam. All associated 95 per cent confidence intervals indicate that the values are statistically significant at the values to be examined.

References

- Baulch, B., and D. Vu (2010). 'Poverty Dynamics in Vietnam, 2002–2006'. Report. Washington, DC: World Bank Available at: <https://openknowledge.worldbank.org/handle/10986/27728> (accessed 15 October 2018).
- Benjamin, D., B. McCaig, L. Brandt, and B. McCaig (2016). 'Growth with Equity: Income Inequality in Vietnam, 2002 – 14'. *The Journal of Economic Inequality*, 15(10392): 25–46.
- Besim, M., and G.P. Jenkins (2005). 'Tax Compliance: When do Employees Behave Like the Self-Employed?'. *Applied Economics*, 37(10): 1201–08.
- Bird, R.M., and E.M. Zolt (2014). 'Redistribution Via Taxation: The Limited Role of the Personal Income Tax in Developing Countries'. *Annals of Economics and Finance*, 15(2): 625–83.
- Castel, P., and T.T. To (2012). 'Informal Employment in the Formal Sector: Wages and Social Security tax Evasion in Vietnam'. *Journal of the Asia Pacific Economy*, 17(4): 616–31.
- Davutyan, N. (2008). 'Estimating the Size of Turkey's Informal Sector: An Expenditure-Based Approach'. *Journal of Economic Policy Reform*, 11(4): 261–71.
- Engström, P., and J. Hagen (2017). 'Income Underreporting Among the Self-Employed: A Permanent Income Approach'. *European Economic Review*, 92: 92–109.
- Engström, P., and B. Holmlund (2009). 'Tax Evasion and Self-Employment in a High-Tax Country: Evidence From Sweden'. *Applied Economics*, 41(19): 2419–30.
- Feldman, N.E., and J. Slemrod (2007). 'Estimating Tax Noncompliance with Evidence from Unaudited Tax Returns'. *Economic Journal*, 117(518): 327–52.
- Giang, D.H., P.T. Xuan, and N.Q. Hai (2016). 'Corruption Risks in Vietnam's Household Business Sector'. *Crime, Law and Social Change*, 65(4–5): 395–422.
- Hagenaars, A., K. de Vos, and M. Zaidi (1994). *Poverty Statistics in the Late 1980s: Research Based on Micro-Data*. Luxembourg: Statistical Office of the European Communities.
- Hansen, H., and T.-L. Dang (2013). 'The Representativeness and Impact of the Vietnam Household Living Standards Surveys'. Copenhagen: Department of Economics, University of Copenhagen.
- Hurst, E., and B. Pugsley (2010). 'Are Household Surveys Like Tax Forms? Evidence from the Self Employed'. *Review of Economics and Statistics*, 96(1): 19–33.
- Johansson, E. (2005). 'An Estimate of Self-Employment Income Underreporting in Finland'. *Nordic Journal of Political Economy*, 31(2): 99–109.
- Kerkvliet, B.J.T. (1995). 'Village–State Relations in Vietnam: The Effect of Everyday Politics on Decollectivization'. *The Journal of Asian Studies*, 54(2): 396–418.
- Kim, B., J. Gibson, and C. Chung (2017). 'Using Panel Data to Estimate Income Under-Reporting by the Self-Employed'. *Manchester School*, 85(1): 41–64.
- Kukk, M., and K. 1 (2014). 'Income Underreporting by Households with Business Income: Evidence From Estonia'. *Post-Communist Economics*, 26(2): 257–76.
- Kukk, M., and K. Staehr (2017). 'Identification of Households Prone to Income Underreporting: Employment Status or Reported Business Income?' *Public Finance Review*, 45(5): 599–627.

- Li, J., and H.A. La (2018). 'Impact of Policy, Market Income Volatility and Demography on Income Inequality in Australia Between 2002 and 2016'. In *IMA Asia-Pacific Conference*, 21–23 March 2018, Narita, Japan.
- Lyssiotou, P., P. Pashardes, and T. Stengos (2004). 'Estimates of the Black Economy Based on Consumer Demand Approaches'. *Economic Journal*, 114(497): 622–40.
- Martinez-Lopez, D. (2013). 'The Underreporting of Income by Self-Employed Workers in Spain'. *SERIEs*, 4(4): 353–71.
- Nguyen, C. (2016). 'The Impact of Health Insurance Programs for Children: Evidence from Vietnam'. *Health Economics Review*, 6(1): 1–15.
- Nguyen, C.H., and H. Liu (2014). 'A Comparative Study on Individual Income Tax Burden of Vietnam and China'. *Business and Management Research*, 3(2): 60–66.
- Nguyen, C.V., and A. Tran (2014). 'Poverty Identification: Practice and Policy Implications in Vietnam'. *Asian-Pacific Economic Literature*, 28(1): 116–36.
- Nguyen, N.A., Q.H. Doan, and B. Tran-Nam (2017). 'Tax Corruption and Private Sector Development in Vietnam'. *eJournal of Tax Research*, 15(2): 290–311.
- OECD. (2014). *Social Cohesion Policy Review of Viet Nam*. Paris: OECD Publishing.
- Painter, M. (2003). 'Public Administration Reform in Vietnam: Problems and Prospects'. *Public Administration and Development*, 23(3): 259–71.
- Pissarides, C.A., and G. Weber (1989). 'An Expenditure-Based Estimate of Britain's Black Economy'. *Journal of Public Economics*, 39(1): 17–32.
- Rand, J., and F. Tarp (2012). 'Firm-Level Corruption in Vietnam'. *Economic Development and Cultural Change*, 60(3): 571–95.
- Schuetze, H.J. (2002). 'Profiles of Tax Non-Compliance Among the Self-Employed in Canada: 1969 to 1992'. *Canadian Public Policy*, 28(2), 219–238.
- Shukla, G.P., D.M. Pham, M. Engelchalk, and T.M. Le (2014). *Tax Reform in Vietnam: Toward a More Efficient and Equitable System*. Washington, DC: World Bank.
- Slemrod, J. (2007). 'Cheating Ourselves: The Economics of Tax Evasion'. *Journal of Economic Perspectives*, 21(1): 25–48.
- Staehr, K. (2004). 'Economic Transition in Estonia: Background, Reforms and Results'. In R. Egle (ed.), *Contemporary Change in Estonia*. Huddinge: Södertörn University College.
- van de Walle, D. (2004). *The Static and Dynamic Incidence of Vietnam's Public Safety Net*. Washington, DC: World Bank.

Appendix A: list of variables

- Total family income including wages and salaries, self-employment income from farming and non-farm production, rental income, investment income, pensions, remittances, and other non-taxable income.
- Total food expenditure including both self-made and purchased food, and eating outside the home.
- Household characteristics:
 - age of household head;
 - dummies of ethnic majority (= 1 if household heads are Vietnamese or Chinese and 0 if heads belong to one of 52 remaining ethnic minorities);
 - working status of household heads (dummies of unemployed heads, heads working at least 30 hours per week (full-time workers), heads working 15–29 hours per week (part-time workers), heads working fewer than 15 hours per week (baseline));
 - number of children aged under 2;
 - number of children aged 2–5;
 - number of children aged 6–10;
 - number of children aged 11–15;
 - number of adults aged 16+;
 - number of household members working at least 15 hours per week.
- Dummies of urban areas (= 1 if households are living in urban areas and 0 otherwise) and regions (dummies of central highlands, Mekong River delta, midlands and northern mountainous areas, northern and coastal central region, Red River delta, and south-east; Red River delta = the baseline).
- Year dummies (= 1 if year = 2014 and 0 if year = 2012).